

GenCore version 5.1.4_p5_4578
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: May 8, 2003, 16:41:13 ; Search time 22 Seconds
(without alignments)
2119.327 Million cell updates/sec

Title: US-09-926-163B-2
Perfect score: 485
Sequence: 1 MITRETLKSLPANVQAPPD.....KVIIVLRKIIREKGVKAAIPA 485

Scoring table: OLIGO
Gapop 60.0 , Gapext 60.0

Searched: 283224 seqs, 96134422 residues

Word size : 0
Total number of hits satisfying chosen parameters: 283224

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Database : PIR_73:*
1: pir1:*
2: pir2:*
3: pir3:*
4: pir4:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	13	2.7	494	2 AG3103	mannotol 2-dehydro
2	13	2.7	525	2 C98183	mannotol 2-dehydro
3	10	2.1	486	2 S56548	fructuronate reduc
4	10	2.1	486	2 B91289	D-mannionate oxidor
5	10	2.1	486	2 E86130	mannonate oxidoreg
6	10	2.1	491	2 D83353	mannotol dehydroge
7	9	1.9	486	2 A64909	probable fructuron
8	9	1.9	486	2 G90897	probable oxidoredu
9	9	1.9	486	2 H85719	probable oxidoredu
10	9	1.9	488	2 AD0679	fructuronate reduc
11	8	1.6	244	2 B76227	hypothetical prote
12	8	1.6	244	2 AD2850	hypothetical prote
13	8	1.6	253	2 S76761	hypothetical prote
14	8	1.6	260	2 I40886	glycine hydroxymet
15	8	1.6	384	2 S73429	glycerol-3-phospat
16	8	1.6	409	2 T36940	probable oxidoredu
17	8	1.6	463	2 AF0283	probable mannotol
18	8	1.6	580	2 A86834	1-deoxyxylulose-5-
19	7	1.4	42	2 T13269	hypothetical prote
20	7	1.4	72	2 H81016	hypothetical prote
21	7	1.4	77	2 A47070	ferrochelatase [EC
22	7	1.4	130	2 A11875	hypothetical prote
23	7	1.4	142	2 S77423	hypothetical prote
24	7	1.4	142	2 C87255	transcriptional regu
25	7	1.4	149	2 A90524	hypothetical prote
26	7	1.4	154	2 B69840	conserved hypothet
27	7	1.4	166	2 D83953	hypothetical prote
28	7	1.4	171	2 C84247	conserved hypothet
29	7	1.4	172	2 T45196	probable transcrip

ALIGNMENTS

RESULT 1

AG3103
mannotol 2-dehydrogenase mtLK [imported] - Agrobacterium tumefaciens (strain C58, Dup
C:Species: Agrobacterium tumefaciens
C:Date: 11-Jan-2002 #sequence_revision 11-Jan-2002 #text_change 01-Feb-2002
C:Accession: AG3103
R:Wood, D.W.; Setubal, J.C.; Kaul, R.; Monks, D.; Chen, L.; Wood, G.E.; Chen, Y.; Woo
erage, G.; Gillet, W.; Grant, C.; Guenther, D.; Kutayavin, T.; Levy, R.; Li, M.; McCl
Science 294, 2317-2323, 2001
A:Authors: Yoo, H.; Tao, Y.; Biddle, P.; Jung, M.; Krespan, W.; Perry, M.; Gordon-Kam
ster, E.W.
A:Title: The Genome of the Natural Genetic Engineer Agrobacterium tumefaciens C58.
A:Reference number: AB2577; PMID:11743193
A:Accession: AG3103
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-494 <KUR>
A:Cross-references: GB:AE008689; PIDN:AA45245.1; PID:g17742929; GSPDB:GN00187
A:Experimental source: strain C58 (Dupont)
C:Genetics:
A:Gene: mtLK
A:Map position: linear chromosome
C:Superfamily: conserved hypothetical protein YEL070w

Query Match 2.7%; Score 13; DB 2; Length 494;
Best Local Similarity 100.0%; Pred.No. 8.3e-05;
Matches 13; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 123 IRIVSWTITEGGY 135
|||||
DB 124 IRIVSWTITEGGY 136

RESULT 2

C98183
mannotol 2-dehydrogenase (mdh) [imported] - Agrobacterium tumefaciens (strain C58, Ce
C:Species: Agrobacterium tumefaciens
C:Date: 22-Oct-2001 #sequence_revision 22-Oct-2001 #text_change 11-Jan-2002
C:Accession: C98183
R:Goodner, B.; Hinkle, G.; Gattung, S.; Miller, N.; Blanchard, M.; Quorollo, B.; Goldm
A.; Liu, F.; Wollam, C.; Allinger, M.; Doughty, D.; Scott, C.; Lappas, C.; Markelz,
Science 294, 2323-2328, 2001
A:Title: Genome Sequence of the Plant Pathogen and Biotechnology Agent Agrobacterium
A:Reference number: A97359; PMID:11743194
A:Accession: C98183
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-525 <KUR>
A:Cross-references: GB:AE007870; PIDN:AAK88989.1; PID:g15158773; GSPDB:GN00170
C:Genetics:
A:Gene: AGR_L_830

A:Map position: linear chromosome

C:Superfamily: conserved hypothetical protein YEL070w

Query Match 2.7% Score 13; DB 2; Length 525;
Best Local Similarity 100.0%; Pred. No. 8.8e-05;

Matches 13; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 123 IRIVSMITEGGY 135
|||||

Db 155 IRIVSMITEGGY 167

RESULT 3

S56548

fructuronate reductase (EC 1.1.1.57) uxuB - Escherichia coli (strain K-12)

N:Alternate names: D-mannonate oxidoreductase

C:Species: Escherichia coli

C>Date: 28-Oct-1995 #sequence_revision 13-Mar-1997 #text_change 01-Mar-2002

C:Accession: S56548; E65246

R:Burland, V.; Plunkett III, G.; Sofia, H.J.; Daniels, D.L.; Blattner, F.R.

Nucleic Acids Res. 23, 2105-2119, 1995

A:Title: Analysis of the Escherichia coli genome VI: DNA sequence of the region from 92.

A:Reference number: S56314; MUID:95334362; PMID:7610040

A:Accession: S56548

A:Status: nucleic acid sequence not shown; translation not shown

A:Molecule type: DNA

A:Residues: 1-486 <BUR>

A:Cross-references: EMBL:U14003; NID:gl263172; PIDN:AAA97219.1; PID:g537164

A:Note: The nucleotide sequence was submitted to the EMBL Data Library, August 1994.

R:Blattner, F.R.; Plunkett III, G.; Bloch, C.A.; Perna, N.T.; Burland, V.; Riley, M.; Co

.A.; Rose, D.J.; Mau, B.; Shao, Y.

Science 277, 1453-1462, 1997

A:Title: The complete genome sequence of Escherichia coli K-12.

A:Reference number: A64720; MUID:97426617; PMID:9278503

A:Accession: E65246

A:Status: nucleic acid sequence not shown; translation not shown

A:Molecule type: DNA

A:Residues: 1-486 <BLAT>

A:Cross-references: GB:AE000503; GB:U00096; NID:gl790777; PIDN:AAC77279.1; PID:gl790779;

A:Experimental source: strain K-12, substrain MG1655

C:Genetics:

A:Gene: uxuB

C:Function:

A:Description: reversible reaction of D-mannonate and NAD(+) to D-fructuronate and NADH

A:Pathway: D-glucuronate catabolism

A:Note: part of the interconversion of D-fructuronate to D-2-keto-3-deoxy-D-gluconate

A:Note: reduces also D-tagaturonate

C:Superfamily: conserved hypothetical protein YEL070w

C:Keywords: carbon-oxygen lyase; hydro-lyase; NAD; oxidoreductase

Query Match 2.1% Score 10; DB 2; Length 486;

Best Local Similarity 100.0%; Pred. No. 0.11;

Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 181 KAFTVMSCDN 190

|||||

Db 178 KAFTVMSCDN 187

RESULT 4

B91289

D-mannonate oxidoreductase EC5282 [imported] - Escherichia coli (strain O157:H7, subst

C:Species: Escherichia coli

C>Date: 18-Jul-2001 #sequence_revision 18-Jul-2001 #text_change 03-Aug-2001

C:Accession: B91289

R:Hayashi, T.; Makino, K.; Ohnishi, M.; Kurokawa, K.; Ishii, K.; Yokoyama, K.; Han, C.G.

gasawara, N.; Yasunaga, T.; Kuhara, S.; Shiba, T.; Hattori, M.; Shinagawa, H.

DNA Res. 8, 11-22, 2001

A:Title: Complete genome sequence of enterohemorrhagic Escherichia coli O157:H7 and genc

A:Reference number: A99629; MUID:21156231; PMID:11258796

A:Accession: B91289

A:Status: preliminary

A:Molecule type: DNA

A:Residues: 1-486 <HAY>

A:Cross-references: GB:BA000007; PIDN:BA38705.1; PID:gl3364760; GSPDB:GN00154

A:Experimental source: strain O157:H7, substrain RIMD 0509952

C:Genetics:

A:Gene: EC5282

C:Superfamily: conserved hypothetical protein YEL070w

Query Match 2.1% Score 10; DB 2; Length 486;

Best Local Similarity 100.0%; Pred. No. 0.11;

Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 181 KAFTVMSCDN 190

|||||

Db 178 KAFTVMSCDN 187

RESULT 5

E86130

mannonate oxidoreductase [imported] - Escherichia coli (strain O157:H7, substrain EDL

C:Species: Escherichia coli

C>Date: 16-Feb-2001 #sequence_revision 16-Feb-2001 #text_change 14-Sep-2001

C:Accession: E86130

R:Perna, N.T.; Plunkett III, G.; Burland, V.; Mau, B.; Glasner, J.D.; Rose, D.J.; May

iller, L.; Grotbeck, E.J.; Davis, N.W.; Lim, A.; Dimalanta, E.; Potamouisis, K.; Apoda

Nature 409, 529-533, 2001

A:Title: Genome sequence of enterohemorrhagic Escherichia coli O157:H7.

A:Reference number: A85480; MUID:21074935; PMID:11206551

A:Accession: E86130

A:Status: preliminary

A:Molecule type: DNA

A:Residues: 1-486 <STO>

A:Cross-references: GB:AE005174; NID:gl2519333; PIDN:AAG59505.1; GSPDB:GN00145; UWGP:

A:Experimental source: strain O157:H7, substrain EDL933

C:Genetics:

A:Gene: uxuB

C:Superfamily: conserved hypothetical protein YEL070w

Query Match 2.1% Score 10; DB 2; Length 486;

Best Local Similarity 100.0%; Pred. No. 0.11;

Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 181 KAFTVMSCDN 190

|||||

Db 178 KAFTVMSCDN 187

RESULT 6

D83353

mannitol dehydrogenase PA2342 [imported] - Pseudomonas aeruginosa (strain PA01)

C:Species: Pseudomonas aeruginosa

C>Date: 15-Sep-2000 #sequence_revision 15-Sep-2000 #text_change 31-Dec-2000

C:Accession: D83353

R:Stover, C.K.; Pham, X.Q.; Erwin, A.L.; Mizoguchi, S.D.; Warren, P.; Hickey, M.J.;

adman, S.; Yuan, Y.; Brody, L.L.; Coulter, S.N.; Folger, K.R.; Kas, A.; Larbig, K.; L

.; Lory, S.; Olson, M.V.

Nature 406, 959-964, 2000

A:Title: Complete genome sequence of Pseudomonas aeruginosa PA01, an opportunistic pa

A:Reference number: A82950; MUID:20437337; PMID:10984043

A:Accession: D83353

A:Status: preliminary

A:Molecule type: DNA

A:Residues: 1-491 <STO>

A:Cross-references: GB:AE004660; GB:AE004091; NID:g9948372; PIDN:AAG05730.1; GSPDB:GN

A:Experimental source: strain PA01

C:Genetics:

A:Gene: mtd; PA2342

C:Superfamily: conserved hypothetical protein YEL070w

Query Match 2.1% Score 10; DB 2; Length 491;

Best Local Similarity 100.0%; Pred. No. 0.11;

Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 182 AFTVMSCDNL 191

Db 183 AFTVMSCDNL 192
|||||

RESULT 7
A64909

probable fructuronate reductase (EC 1.1.1.57) ydfI - Escherichia coli (strain K-12)
N:Alternate names: D-mannionate oxidoreductase
C:Species: Escherichia coli
C:Date: 12-Sep-1997 #sequence_revision 17-Sep-1997 #text_change 01-Mar-2002
C:Accession: A64909
R:Blattner, F.R.; Plunkett III, G.; Bloch, C.A.; Perna, N.T.; Burland, V.; Riley, M.; Cohen, A.; Rose, D.J.; Mau, B.; Shao, Y.
Science 277, 1453-1462, 1997
A:Title: The complete genome sequence of Escherichia coli K-12.
A:Reference number: A64720; MUID:97426617; PMID:9278503
A:Accession: A64909
A:Status: nucleic acid sequence not shown; translation not shown
A:Molecule type: DNA
A:Residues: 1-486 <BLAT>
A:Cross-references: GB:AB000251; GB:U00096; NID:g1787814; PIDN:AAC74615.1; PID:g1787823;
A:Experimental source: strain K-12, substrain MGL655
C:Genetics:
A:Gene: ydfI
C:Function:
A:Description: reversible reaction of D-mannionate and NAD(+) to D-fructuronate and NADH
A:Pathway: D-glucuronate catabolism
A:Note: part of the interconversion of D-fructuronate to D-2-keto-3-deoxy-D-gluconate
A:Note: reduces also D-tagaturonate
C:Superfamily: conserved hypothetical protein YEL070w
C:Keywords: carbon-oxygen lyase; hydro-lyase; NAD; oxidoreductase

Query Match 1.9%; Score 9; DB 2; Length 486;
Best Local Similarity 100.0%; Pred. No. 1.2;
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 182 AFTVMSCDN 190
|||||

Db 176 AFTVMSCDN 184

RESULT 8
G90897

probable oxidoreductase ECs2151 [imported] - Escherichia coli (strain O157:H7, substrain K-12)
C:Species: Escherichia coli
C:Date: 18-Jul-2001 #sequence_revision 18-Jul-2001 #text_change 03-Aug-2001
C:Accession: G90897
R:Hayashi, T.; Makino, K.; Ohnishi, M.; Kurokawa, K.; Ishii, K.; Yokoyama, K.; Han, C.G.; Gasawara, N.; Yasunaga, T.; Kuhara, S.; Shiba, T.; Hattori, M.; Shinagawa, H.
DNA Res. 8, 11-22, 2001
A:Title: Complete genome sequence of enterohemorrhagic Escherichia coli O157:H7 and genomic islands
A:Reference number: A99629; MUID:21156231; PMID:11258796
A:Accession: G90897
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-486 <HAY>
A:Cross-references: PIDN:BA000007; PIDN:BA035574.1; PID:g13361617; GSPDB:GN00154
A:Experimental source: strain O157:H7, substrain RIMD 0509952
C:Genetics:
A:Gene: ECs2151
C:Superfamily: conserved hypothetical protein YEL070w

Query Match 1.9%; Score 9; DB 2; Length 486;
Best Local Similarity 100.0%; Pred. No. 1.2;
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 182 AFTVMSCDN 190
|||||

Db 176 AFTVMSCDN 184

RESULT 9
H85719

probable oxidoreductase ydfI [imported] - Escherichia coli (strain O157:H7, substrain K-12)
C:Species: Escherichia coli
C:Date: 16-Feb-2001 #sequence_revision 16-Feb-2001 #text_change 14-Sep-2001
C:Accession: H85719
R:Perna, N.T.; Plunkett III, G.; Burland, V.; Mau, B.; Glasner, J.D.; Rose, D.J.; May, A.; Miller, L.; Grobeck, E.J.; Davis, N.W.; Lim, A.; Dimalanta, E.; Potamousis, K.; Apoda, N.; Nature 409, 529-533, 2001
A:Title: Genome sequence of enterohemorrhagic Escherichia coli O157:H7.
A:Reference number: A85480; MUID:21074935; PMID:11206551
A:Accession: H85719
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-486 <STO>
A:Cross-references: NID:g12515111; PIDN:AAG56220.1; GSPDB:GN00145; UWGP:
A:Experimental source: strain O157:H7, substrain EDL933
C:Genetics:
A:Gene: ydfI
C:Superfamily: conserved hypothetical protein YEL070w

Query Match 1.9%; Score 9; DB 2; Length 486;
Best Local Similarity 100.0%; Pred. No. 1.2;
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 182 AFTVMSCDN 190
|||||

Db 176 AFTVMSCDN 184

RESULT 10
AD0679
fructuronate reductase (EC 1.1.1.57) - Salmonella enterica subsp. enterica serovar Typhimurium
C:Species: Salmonella enterica subsp. enterica serovar Typhi
A:Note: this species has also been called Salmonella typhi
C:Date: 09-Nov-2001 #sequence_revision 09-Nov-2001 #text_change 27-Nov-2001
C:Accession: AD0679
R:Parkhill, J.; Dougan, G.; James, K.D.; Thomson, N.R.; Pickard, D.; Wain, J.; Church, T.; Connor, P.; Cronin, A.; Davis, P.; Davies, R.M.; Dowd, L.; White, N.; Farr, S.; Moulle, S.; O'Gaora, P.
Nature 413, 848-852, 2001
A:Authors: Parry, C.; Ouail, M.; Rutherford, K.; Simmonds, M.; Skelton, J.; Stevens, A.; Title: Complete genome sequence of a multiple drug resistant Salmonella enterica serovar Typhimurium
A:Reference number: AB0502; PMID:11677608
A:Accession: AD0679
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-488 <PAR>
A:Cross-references: GB:AL513382; PIDN:CAD01805.1; PID:g16502650; GSPDB:GN00176
C:Genetics:
A:Gene: STY1553
C:Superfamily: conserved hypothetical protein YEL070w
C:Keywords: oxidoreductase

Query Match 1.9%; Score 9; DB 2; Length 488;
Best Local Similarity 100.0%; Pred. No. 1.2;
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 182 AFTVMSCDN 190
|||||

Db 176 AFTVMSCDN 184

RESULT 11
B97627

hypothetical protein AGR_C_4053 [imported] - Agrobacterium tumefaciens (strain C58, C58:G)
C:Species: Agrobacterium tumefaciens
C:Date: 30-Sep-2001 #sequence_revision 30-Sep-2001 #text_change 11-Jan-2002
C:Accession: B97627
R:Goodner, B.; Hinkle, G.; Gattung, S.; Miller, N.; Blanchard, M.; Ouello, B.; Goldm, A.; Liu, F.; Wollam, C.; Allinger, M.; Doughty, D.; Scott, C.; Lappas, C.; Markelz, S.; Science 294, 2323-2328, 2001
A:Title: Genome Sequence of the Plant Pathogen and Biotechnology Agent Agrobacterium tumefaciens
A:Reference number: A97359; PMID:11743194
A:Accession: B97627

A>Status: preliminary
 A:Molecule type: DNA
 A:Residues: 1-244 <KUR>
 A:Cross-references: GB:AE007869; PIDN:AAK87971.1; PID:gl15157379; GSPDB:GN00169
 C:Genetics:
 A:Gene: AGR_C_4053
 A:Map position: circular chromosome

Query Match 1.6%; Score 8; DB 2; Length 244;
 Best Local Similarity 100.0%; Pred. No. 6.9;
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 6 TLKSLPAN 13
 Db 183 TLKSLPAN 190

RESULT 12
 AD2850
 hypothetical protein Atu2229 [Imported] - Agrobacterium tumefaciens (strain C58, Dupont)
 C:Species: Agrobacterium tumefaciens
 C:Date: 11-Jan-2002 #sequence_revision 11-Jan-2002 #text_change 11-Jan-2002
 C:Accession: AD2850
 R:Wood, D.W.; Setubal, J.C.; Kaul, R.; Monks, D.; Chen, L.; Wood, G.E.; Chen, Y.; Woo, I.; Karp, P.; Gillet, W.; Grant, C.; Guenther, D.; Kutyavin, T.; Levy, R.; Li, M.; McClellan, S.; Science 294, 2317-2323, 2001
 A:Authors: Yoo, H.; Tao, Y.; Biddle, P.; Jung, M.; Krespan, W.; Perry, M.; Gordon-Kamm, S.E.W.
 A:Title: The Genome of the Natural Genetic Engineer Agrobacterium tumefaciens C58.
 A:Reference number: AB2577; PMID:11743193
 A:Accession: AD2850
 A>Status: preliminary
 A:Molecule type: DNA
 A:Residues: 1-244 <KUR>
 A:Cross-references: GB:AE008688; PIDN:AAL43218.1; PID:gl17740700; GSPDB:GN00186
 A:Experimental source: strain C58 (Dupont)
 C:Genetics:
 A:Gene: Atu2229
 A:Map position: circular chromosome

Query Match 1.6%; Score 8; DB 2; Length 244;
 Best Local Similarity 100.0%; Pred. No. 6.9;
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 6 TLKSLPAN 13
 Db 183 TLKSLPAN 190

RESULT 13
 S76761
 hypothetical protein - Synecocystis sp. (strain PCC 6803)
 C:Species: Synecocystis sp.
 A:Variety: PCC 6803
 C:Date: 25-Apr-1997 #sequence_revision 25-Apr-1997 #text_change 20-Jun-2000
 R:Kaneko, T.; Sato, S.; Kotani, H.; Tanaka, A.; Asamizu, E.; Nakamura, Y.; Miyajima, N.; O, K.; Okumura, S.; Shimpo, S.; Takeuchi, C.; Wada, T.; Watanabe, A.; Yamada, M.; Yasuda, K.; DNA Res. 3, 109-136, 1996
 A:Title: Sequence analysis of the genome of the unicellular cyanobacterium Synecocystis sp.
 A:Reference number: S74322; MUID:97061201; PMID:8905231
 A:Accession: S76761
 A>Status: nucleic acid sequence not shown; translation not shown
 A:Molecule type: DNA
 A:Residues: 1-253 <KAN>
 A:Cross-references: EMBL:D90916; GB:AB001339; NID:gl653715; PIDN:BAAL18673.1; PID:gl65376
 A:Note: the nucleotide sequence was submitted to the EMBL Data Library, June 1996
 C:Genetics:
 A:Start codon: GTG

Query Match 1.6%; Score 8; DB 2; Length 253;

Best Local Similarity 100.0%; Pred. No. 7.1;
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 338 VIPTLKAP 345
 Db 2 VIPTLKAP 9

RESULT 14
 I40886
 glycine hydroxymethyltransferase (EC 2.1.1.2.1) - Corynebacterium sp. (fragment)
 N:Alternate names: serine aldolase; serine hydroxymethylase; threonine aldolase
 C:Species: Corynebacterium sp.
 C:Date: 16-Aug-1996 #sequence_revision 16-Aug-1996 #text_change 22-Jun-1999
 C:Accession: I40886
 R:Chlumsky, L.J.; Zhang, L.; Jorns, M.S.
 J. Biol. Chem. 270, 18252-18259, 1995
 A:Title: Sequence analysis of sarcosine oxidase and nearby genes reveals homologies w
 A:Reference number: A57385; MUID:95355441; PMID:7543100
 A:Accession: I40886
 A>Status: preliminary; translated from GB/EMBL/DBJ
 A:Molecule type: DNA
 A:Residues: 1-260 <RES>
 A:Cross-references: EMBL:U23955; NID:927587; PIDN:AAC43458.1; PID:927588
 C:Genetics:
 A:Gene: glyA
 C:Superfamily: glycine hydroxymethyltransferase
 C:Keywords: phosphoprotein; pyridoxal phosphate; transferase
 F:60/Binding site: Pyridoxal phosphate (Lys) (covalent) #status predicted

Query Match 1.6%; Score 8; DB 2; Length 260;
 Best Local Similarity 100.0%; Pred. No. 7.3;
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 237 AEIAKKLN 244
 Db 75 AEIAKKLN 82

RESULT 15
 S73429
 glycerol-3-phosphate dehydrogenase glpD, aerob induced - Mycoplasma pneumoniae (strain
 N:Alternate names: hypothetical protein D09_orf384
 C:Species: Mycoplasma pneumoniae
 A:Variety: ATCC 29342
 C:Date: 26-Feb-1997 #sequence_revision 25-Apr-1997 #text_change 19-May-2000
 C:Accession: S73429
 R:Himmelreich, R.; Hilbert, H.; Plagens, H.; Pirkel, E.; Li, B.C.; Herrmann, R.
 Nucleic Acids Res. 24, 4420-4449, 1996
 A:Title: Complete sequence analysis of the genome of the bacterium Mycoplasma pneumonia
 A:Reference number: S73327; MUID:97105885; PMID:8948633
 A:Accession: S73429
 A>Status: preliminary; nucleic acid sequence not shown; translation not shown
 A:Molecule type: DNA
 A:Residues: 1-384 <HIM>
 A:Cross-references: EMBL:AE000012; GB:U00089; NID:gl673755; PIDN:AA895751.1; PID:gl67
 A:Note: the nucleotide sequence was submitted to the EMBL Data Library, November 1996
 C:Genetics:
 A:Gene: glpD
 A:Genetic code: SCS3
 C:Superfamily: Escherichia coli hypothetical protein ygaF

Query Match 1.6%; Score 8; DB 2; Length 384;
 Best Local Similarity 100.0%; Pred. No. 10;
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 437 LAKADDFE 444
 Db 219 LAKADDFE 226

Search completed: May 8, 2003, 16:45:11
 Job time : 34 secs

THIS PAGE BLANK (USPTO)

GenCore version 5.1.4_p5_4578
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: May 8, 2003, 16:35:24 ; Search time 14 Seconds
(without alignments)
1436.859 Million cell updates/sec

Title: US-09-926-163B-2
Perfect score: 485
Sequence: 1 MITRETLKSLPANVOAPPDY.....KVIVLRKIIRKGVKAIPA 485

Scoring table: OLICO

Gapop 60.0 , Gapext 60.0

Searched: 112892 seqs, 41476328 residues

Word size : 0

Total number of hits satisfying chosen parameters: 112892

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Database : SwissProt_40:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	49	10.1	50	POR_GLUOX	P80354 gluconobact
2	11	2.3	476	MTLK_RHOSH	P33216 rhodobacter
3	10	2.1	486	UXUB_ECOLI	P39160 escherichia
4	9	1.9	486	YDFI_ECOLI	P77260 escherichia
5	8	1.6	260	GLYA_CORSI	P50463 corynebacte
6	8	1.6	384	Y039_MYCPN	P75063 mycoplasma
7	8	1.6	463	DAUD_YERPE	P58709 yersinia pe
8	7	1.4	77	HEMZ_YERPS	O05338 yersinia ps
9	7	1.4	142	PTP_ACTJO	O52787 acinetobact
10	7	1.4	171	Y428_MYCGE	P47667 mycoplasma
11	7	1.4	172	Y428_MYCPN	P75169 mycoplasma
12	7	1.4	179	YPIB_BACSU	P54390 bacillus su
13	7	1.4	182	WCAF_ECOLI	P71240 escherichia
14	7	1.4	200	DGC6_CHICK	O73770 gallus gall
15	7	1.4	200	GTT1_DROER	P30104 drosophila
16	7	1.4	200	GTT1_DROMA	P30105 drosophila
17	7	1.4	200	GTT1_DROSE	P30106 drosophila
18	7	1.4	200	GTT1_DROTE	P30107 drosophila
19	7	1.4	200	GTT1_DROVA	P30108 drosophila
20	7	1.4	203	RPOC_PROSC	P42078 prochloron
21	7	1.4	209	GTT1_DROME	P20432 drosophila
22	7	1.4	210	RLI1_CHICK	O08200 gallus gall
23	7	1.4	211	CCGX_MOUSE	O94jv3 mus musculu
24	7	1.4	216	RCSB_ECOLI	P14374 escherichia
25	7	1.4	216	RCSB_SALTI	O56127 salmonella
26	7	1.4	216	RCSB_SALTY	P58663 salmonella
27	7	1.4	232	YAGL_ECOLI	P77607 escherichia
28	7	1.4	248	TPIS_PLAFA	O07412 plasmodium
29	7	1.4	280	VG85_BPPH2	P20344 bacterioph
30	7	1.4	299	YFAD_ECOLI	P37014 escherichia
31	7	1.4	303	VANY_ENTFC	P37711 enterococu
32	7	1.4	303	YDIP_ECOLI	P77402 escherichia
33	7	1.4	308	CPP3_ENTHI	O06964 entamoeba h

RESULT 1				
ID	POR_GLUOX	STANDARD;	PRT;	50 AA.
AC	P80354;			
DT	01-NOV-1995 (Rel. 32, Created)			
DT	01-NOV-1995 (Rel. 32, Last sequence update)			
DT	16-OCT-2001 (Rel. 40, Last annotation update)			
DE	Polyol:NADP oxidoreductase (EC 1.1.1.-) (fragment).			
POR.				
OS	Gluconobacter oxydans (Gluconobacter suboxydans).			
OC	Bacteria; Proteobacteria; alpha subdivision; Acetobacteraceae;			
OC	Gluconobacter.			
OX	NCBI_TaxID=442;			
RN	[1]			
RP	SEQUENCE.			
RC	STRAIN=DSM 350;			
RA	Klasen R.;			
RL	Thesis (1994), Heinrich-Heine University / Duesseidorf, Germany.			
CC	!- SUBCELLULAR LOCATION: Cytoplasmic.			
DR	InterPro: IPR000669; Mannitol_dh.			
DR	Pfam: PF01232; Mannitol_dh; 1.			
KW	Oxidoreductase; NADP.			
FT	NON_TER 50 50			
SQ	SEQUENCE 50 AA; 5648 MW; 8EE7F5B6413497FD CRC64;			
Query Match				
Best Local Similarity 100.0%; Pred. No. 4.2e-42;				
Matches 49; Conservative 0; Mismatches 0; Indels 0;				
QY	1	MITRETLKSLPANVOAPPDYDIDGIRKPGIVHFGVGNFFRAHEAFYVEQIL	49	
Db	1	MITRETLKSLPANVOAPPDYDIDGIRKPGIVHFGVGNFFRAHEAFYVEQIL	49	
RESULT 2				
ID	MTLK_RHOSH	STANDARD;	PRT;	476 AA.
AC	P33216;			
DT	01-FEB-1994 (Rel. 28, Created)			
DT	01-OCT-1994 (Rel. 30, Last sequence update)			
DT	15-JUL-1998 (Rel. 36, Last annotation update)			
DE	Mannitol 2-dehydrogenase (EC 1.1.1.67) (MDH).			
GN	MTLK.			
OS	Rhodobacter sphaeroides (Rhodopseudomonas sphaeroides).			
OC	Bacteria; Proteobacteria; alpha subdivision; Rhodobacter group;			
OC	Rhodobacter.			
OX	NCBI_TaxID=1063;			
RN	[1]			
RP	SEQUENCE FROM N.A., AND SEQUENCE OF 1-43.			
RC	STRAIN=SI-4;			
RX	MEDLINE=94075966; PubMed=8254318;			
RA	Schneider K.-H., Giffhorn F., Kaplan S.;			
RT	"Cloning, nucleotide sequence and characterization of the mannitol			
RT	dehydrogenase gene from Rhodobacter sphaeroides.";			
RL	J. Gen. Microbiol. 139:2475-2484(1993).			
RN	[2]			

P14295 lactobacill
Q8zc98 yersinia pe
P43413 yersinia en
Q8xb0 ralstonia s
P51091 maus domes
Q07104 mus musculu
Q30570 brucella ab
Q99726 homo sapien
P97441 mus musculu
P50855 actinobacil
Q17103 asterias vu
P24607 pseudomonas

```
RP SEQUENCE FROM N.A.
RC STRAIN-SI-4;
RA Schneider K.-H., Giffhorn F.;
RL Submitted (MAR-1994) to the EMBL/GenBank/DBJ databases.
CC -!- CATALYTIC ACTIVITY: D-mannitol + NAD(+) = D-fructose + NADH.
CC -!- SUBUNIT: MONOMER.
CC -!- SIMILARITY: BELONGS TO THE MANNITOL DEHYDROGENASES FAMILY.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announcement/
CC or send an email to license@isb-sib.ch).
CC
CC EMBL; AF018073; AAC45771.1; -.
DR InterPro; IPR000669; Mannitol_dh.
DR Pfam; PF01232; Mannitol_dh; 1.
DR PRINTS; PR00084; MTLDHDRGNASE.
DR PROSITE; PS00974; MANNITOL_DHGENASE; 1.
KW Oxidoreductase; NAD.
FT INIT-MET 0
FT NP_BIND 18 29 NAD (BY SIMILARITY).
SQ SEQUENCE 476 AA; 51406 MW; A58247D6C7CCAF29 CRC64;
Query Match 2.3%; Score 11; DB 1; Length 476;
Best Local Similarity 100.0%; Pred. No. 0.0064;
Matches 11; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 223 FPNQVMDRITP 233
DB 211 FPNQVMDRITP 221
RESULT 3
UXUB_ECOLI
ID UXUB_ECOLI STANDARD; PRT; 486 AA.
AC P39160;
DT 01-FEB-1995 (Rel. 31, Created)
DT 01-FEB-1995 (Rel. 31, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE D-mannionate oxidoreductase (EC 1.1.1.57) (Fructuronate reductase).
GN UXUB OR B4323.
OS Escherichia coli.
OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
OC Escherichia.
OX NCBI_TaxID=562;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-K12 / MG1655;
RX MEDLINE=95334362; PubMed=7610040;
RA Burland V.D., Plunkett G. III, Sofia H.J., Daniels D.L.,
RA Blattner F.R.;
RT "Analysis of the Escherichia coli genome VI: DNA sequence of the
RT region from 92.8 through 100 minutes.";
RL Nucleic Acids Res. 23:2105-2119(1995).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN-K12 / W3110;
RA Mizobuchi K.;
RL Submitted (SEP-1992) to the EMBL/GenBank/DBJ databases.
CC -!- CATALYTIC ACTIVITY: D-mannonate + NAD(+) = D-fructuronate + NADH.
CC -!- SIMILARITY: BELONGS TO THE MANNITOL DEHYDROGENASES FAMILY.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announcement/
CC or send an email to license@isb-sib.ch).
CC
```

```
DR EMBL; U14003; AAA97219.1; -.
DR EMBL; AE000503; AAC77279.1; -.
DR EMBL; D13329; BAA02591.1; -.
DR EcoGene; EG20248; uxuB.
DR InterPro; IPR000669; Mannitol_dh.
DR Pfam; PF01232; Mannitol_dh; 1.
DR PRINTS; PR00084; MTLDHDRGNASE.
DR PROSITE; PS00974; MANNITOL_DHGENASE; 1.
KW Oxidoreductase; NAD; Complete proteome.
FT NP_BIND 25 36 NAD (BY SIMILARITY).
SQ SEQUENCE 486 AA; 5380 MW; 7835638682AF7B87 CRC64;
Query Match 2.1%; Score 10; DB 1; Length 486;
Best Local Similarity 100.0%; Pred. No. 0.067;
Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 181 KAFVTMSCDN 190
DB 178 KAFVTMSCDN 187
RESULT 4
YDFI_ECOLI
ID YDFI_ECOLI STANDARD; PRT; 486 AA.
AC P77260;
DT 01-NOV-1997 (Rel. 35, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Hypothetical oxidoreductase ydfI (EC 1.-.-.-).
GN YDFI OR B1542.
OS Escherichia coli.
OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
OC Escherichia.
OX NCBI_TaxID=562;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-K12 / MG1655;
RX MEDLINE=97426617; PubMed=9278503;
RA Blattner F.R., Plunkett G. III, Bloch C.A., Perna N.T., Burland V.,
RA Riley M., Collado-Vides J., Glasner J.D., Rode C.K., Mayhew G.F.,
RA Gregor J., Davis N.W., Kirkpatrick H.A., Goeden M.A., Rose D.J.,
RA Mau B., Shao Y.;
RT "The complete genome sequence of Escherichia coli K-12.";
RL Science 277:1453-1474(1997).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN-K12;
RX MEDLINE=97251357; PubMed=9097039;
RA Aliba H., Baba T., Fujita K., Hayashi K., Inada T., Isono K.,
RA Itoh T., Kasai H., Kashimoto K., Kimura S., Kitakawa M.,
RA Kitagawa M., Makino K., Miki T., Mizobuchi K., Mori H., Mori T.,
RA Motomura K., Nakade S., Nakamura Y., Nishimoto H., Nishio Y.,
RA Oshima T., Saito N., Sampei G., Seki Y., Sivasubram S.,
RA Tagami H., Takeda J., Takemoto K., Takeuchi Y., Wada C.,
RA Yamamoto Y., Horiuchi T.;
RT "A 570-kb DNA sequence of the Escherichia coli K-12 genome
RT corresponding to the 28.0-40.1 min region on the linkage map.";
RL DNA Res. 3:363-377(1996).
CC -!- SIMILARITY: BELONGS TO THE MANNITOL DEHYDROGENASES FAMILY.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announcement/
CC or send an email to license@isb-sib.ch).
CC
CC EMBL; AE000251; AAC74615.1; -.
DR EMBL; D90797; BAA15243.1; -.
DR EMBL; D90798; BAA15247.1; -.
DR EcoGene; EG13821; ydfI.
DR InterPro; IPR000669; Mannitol_dh.
```

```

DR PFAM: PF01232; Mannitol_dh; 1.
DR PRINTS: PR00084; MTLHDHGNASE.
DR PROSITE: PS00974; MANNITOL_DHGENASE; 1.
KW Hypothetical protein; Oxidoreductase; NAD; Complete proteome.
FT NP_BIND 24 35 NAD (BY SIMILARITY).
SQ SEQUENCE 486 AA; 53685 MW; 96F3BF07AE114B70 CRC64;

Query Match 1.9%; Score 9; DB 1; Length 486;
Best Local Similarity 100.0%; Pred. No. 0.68;
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 182 AFTVMSCDN 190
DB 176 AFTVMSCDN 184
|||||
* RESULT 5
GLYA_CORS1 STANDARD; PRT; 260 AA.
AC P50434;
DT 01-OCT-1996 (Rel. 34, Created)
DT 01-OCT-1996 (Rel. 34, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Serine hydroxymethyltransferase (EC 2.1.2.1) (Serine methylase)
DE (SHMT) (Fragment).
GN GLYA.
OS Corynebacterium sp. (strain P-1).
OC Bacteria; Actinobacteria; Actinobacteriia (class); Actinobacteridae;
OC Actinomycetales; Corynebacterineae; Corynebacteriaceae;
OC Corynebacterium.
OX NCBI_TaxID=69006;
RN [1]
SEQUENCE FROM N.A.
RP MEDLINE=9535441; PubMed=7543100;
RX Chlumsky L.J., Zhang L., Jorns M.S.;
RA "Sequence analysis of sarcosine oxidase and nearby genes reveals
RT homologies with key enzymes of folate one-carbon metabolism.";
RL J. Biol. Chem. 270:18252-18259(1995).
CC -!- FUNCTION: INTERCONVERSION OF SERINE AND GLYCINE.
CC H(2)O = tetrahydrofolate + L-serine.
CC -!- CATALYTIC ACTIVITY: 5,10-methylenetetrahydrofolate + glycine +
CC -!- COPACTOR: PYRIDOXAL PHOSPHATE (BY SIMILARITY).
CC -!- PATHWAY: KEY ENZYME IN THE BIOSYNTHESIS OF PURINES, LIPIDS,
CC HORMONES AND OTHER COMPONENTS.
CC -!- SUBUNIT: HOMOTETRAMER (BY SIMILARITY).
CC -!- SUBCELLULAR LOCATION: Cytoplasmic.
CC -!- SIMILARITY: BELONGS TO THE SHMT FAMILY.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
DR EMBL; U23955; AAC43458.1; -
DR HSSP; P07511; 1C70.
DR InterPro; IPR001085; Gly_HyMettransf.
DR PFAM: PF00464; SHMT; 1.
DR PROSITE; PS00096; SHMT; 1.
KW Nontransferase; Pyridoxal phosphate; One-carbon metabolism.
FT NON_TER 1 1
FT BINDING 60 60 PYRIDOXAL PHOSPHATE (BY SIMILARITY).
SQ SEQUENCE 260 AA; 27578 MW; AD16613C7DAE467B CRC64;

Query Match 1.6%; Score 8; DB 1; Length 260;
Best Local Similarity 100.0%; Pred. No. 4;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 237 AEIAKKLN 244
DB 75 AEIAKKLN 82
|||||
us-09-926-163b-2.rsp
* RESULT 6
Y039_MYCPN STANDARD; PRT; 384 AA.
AC P75063;
DT 01-NOV-1997 (Rel. 35, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Hypothetical protein MG039 homolog (D09_orf384).
GN MPN051 OR MP103.
OS Mycoplasma pneumoniae.
OC Bacteria; Firmicutes; Mollicutes; Mycoplasmataceae; Mycoplasma.
OX NCBI_TaxID=2104;
RN [1]
SEQUENCE FROM N.A.
RP STRAIN=ATCC 29342 / M129;
RX MEDLINE=97105885; PubMed=8948633;
RA Himmelreich R., Hilbert H., Plagens H., Pirkel E., Li B.-C.,
RA Herrmann R.;
RT "Complete sequence analysis of the genome of the bacterium Mycoplasma
RL pneumoniae.";
CC Nucleic Acids Res. 24:4420-4449(1996).
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
DR EMBL; AE000012; AAB95751.1; -
DR InterPro; IPR000927; D_aa_oxidase.
DR PFAM: PF01266; DAO; 1.
KW Hypothetical protein; Complete proteome.
SQ SEQUENCE 384 AA; 42724 MW; DA3E128719EE1BCD CRC64;

Query Match 1.6%; Score 8; DB 1; Length 384;
Best Local Similarity 100.0%; Pred. No. 5.6;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 437 LAKADDFE 444
DB 219 LAKADDFE 226
|||||
* RESULT 7
DALD_YERPE STANDARD; PRT; 463 AA.
AC P58709;
DT 15-JUN-2002 (Rel. 41, Created)
DT 15-JUN-2002 (Rel. 41, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE D-arabinitol 4-dehydrogenase (EC 1.1.1.11).
GN DALD OR YPO2325.
OS Yersinia pestis.
OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
OC Yersinia.
OX NCBI_TaxID=632;
RN [1]
SEQUENCE FROM N.A.
RP STRAIN=CO-92 / Biovar Orientalis;
RX MEDLINE=21470413; PubMed=11586360;
RA Parkhill J., Wren B.W., Thomson N.R., Titball R.W., Holden M.T.G.,
RA Prentice M.B., Sebahia M., James K.D., Churcher C., Mungall K.L.,
RA Baker S., Basham D., Bentley S.D., Brooks K., Cerdeno-Tarraga A.M.,
RA Chillingworth T., Cronin A., Davies R.M., Davis P., Dougan G.,
RA Feltwell T., Hamlin N., Holroyd S., Jagels K., Karlyshev A.V.,
RA Leather S., Moule S., Oyston P.C.F., Quail M., Rutherford K.,
RA Simmonds M., Skelton J., Stevens K., Whitehead S., Barrell B.G.;
RT "Genome sequence of Yersinia pestis, the causative agent of plague.";
RL Nature 413:523-527(2001).

```



```
QY 239 IAKKLN 245
DB 65 IAKKLN 71
|||||
RESULTS
Y428_MYCSE STANDARD; PRT; 171 AA.
AC P47667;
DT 01-FEB-1996 (Rel. 33, Created)
DT 01-FEB-1996 (Rel. 33, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Hypothetical protein MG428.
GN MG428.
OS Mycoplasma genitalium.
OC Bacteria; Firmicutes; Mollicutes; Mycoplasmataceae; Mycoplasma.
ON NCBI_TaxID=2097;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-ATCC 33530 / G-37;
RX MEDLINE=96026346; PubMed=7569993;
RA Fraser C.M., Gocayne J.D., White O., Adams M.D., Clayton R.A.,
RA Fleischmann R.D., Bult C.J., Kerlavage A.R., Sutton G., Kelley J.M.,
RA Fritchman J.L., Weidman J.F., Small K.V., Sandusky M., Fuhrmann J.L.,
RA Nguyen D.T., Otterback T.R., Saudek D.M., Phillips C.A., Merrick J.M.,
RA Tomb J.-F., Dougherty B.A., Bott K.F., Hu P.-C., Lucier T.S.,
RA Peterson S.N., Smith H.O., Hutchison C.A. III, Venter J.C.;
RT "The minimal gene complement of Mycoplasma genitalium";
RL Science 270:397-403(1995).
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; U39725; AAC72449.1; -
DR TIGR; MG428; -
DR InterPro; IPR000792; HTH_LuxR.
KW Hypothetical protein; Complete proteome.
SQ SEQUENCE 171 AA; 20259 MW; 87FB346C3345E5E72 CRC64;
Query Match 1.4%; Score 7; DB 1; Length 171;
Best Local Similarity 100.0%; Pred. No. 28;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 238 ETAKLN 244
DB 139 ETAKLN 145
|||||
RESULTS
Y428_MYCSE STANDARD; PRT; 172 AA.
AC P75169;
DT 01-NOV-1997 (Rel. 35, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Hypothetical protein MG428 homolog (Cl2_orf172).
GN MPN026 OR MP216.
OS Mycoplasma pneumoniae.
OC Bacteria; Firmicutes; Mollicutes; Mycoplasmataceae; Mycoplasma.
ON NCBI_TaxID=2104;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-ATCC 29342 / M129;
RX MEDLINE=97105985; PubMed=8948633;
RA Himmelreich R., Hilbert H., Plagens H., Pirkel E., Li B.-C.,
RA Herrmann R.;
RT "Complete sequence analysis of the genome of the bacterium Mycoplasma
pneumoniae.";
RL Nucleic Acids Res. 24:4420-4449(1996).
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; AF000021; AAB95864.1; -
DR InterPro; IPR000792; HTH_LuxR.
KW Hypothetical protein; Complete proteome.
SQ SEQUENCE 172 AA; 20473 MW; 93F6430F3BF8C4AB CRC64;
Query Match 1.4%; Score 7; DB 1; Length 172;
Best Local Similarity 100.0%; Pred. No. 28;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 238 ETAKLN 244
DB 140 ETAKLN 146
|||||
RESULTS
YPIB_BACSU STANDARD; PRT; 179 AA.
AC P54390;
DT 01-OCT-1996 (Rel. 34, Created)
DT 01-OCT-1996 (Rel. 34, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Hypothetical protein ypiB.
GN YPIB.
OS Bacillus subtilis.
OC Bacteria; Firmicutes; Bacillales; Bacillaceae; Bacillus.
ON NCBI_TaxID=1423;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=168;
RX STRAIN=168 / Marburg;
RX MEDLINE=96349105; PubMed=8760912;
RA Sorokin A.V., Azevedo V., Zumstein E., Galleron N., Ehrlich S.D.,
RA Serror P.;
RT "Sequence analysis of the Bacillus subtilis chromosome region between
RT the serA and kdg loci cloned in a yeast artificial chromosome.";
RL Microbiology 142:2005-2016(1996).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=168;
RX MEDLINE=98044033; PubMed=9384377;
RA Kunst F., Ogasawara N., Moszer I., Albertini A.M., Alloni G.,
RA Azevedo V., Bertero M.G., Bessieres P., Bolotin A., Borchert S.,
RA Borriess R., Boursier L., Brans A., Braun M., Brignell S.C., Bron S.,
RA Brouillet S., Bruschi C.V., Caldwell B., Capuano V., Carter N.M.,
RA Choi S.K., Codani J.J., Connerton I.F., Cummings N.J., Daniel R.A.,
RA Denizot F., Devine K.M., Dusterhoft A., Ehrlich S.D., Emerson P.T.,
RA Entian K.D., Errington J., Fabret C., Ferrari E., Foulger D.,
RA Fritz C., Fujita M., Fujita Y., Fuma S., Galizzi A., Galleron N.,
RA Ghim S.Y., Glaser P., Goffeau A., Gollightly E.J., Grandi G.,
RA Guiseppi G., Guy B.J., Haga K., Haiech J., Harwood C.R., Henaut A.,
RA Hilbert H., Holsappel S., Hosono S., Hulio M.F., Itaya M., Jones L.,
RA Joris B., Karamata D., Kasahara Y., Klaerr-Blanchard M., Klein C.,
RA Kobayashi Y., Koetter P., Koningsstein G., Krogh S., Kumano M.,
RA Kurita K., Lapidus A., Lardinol S., Lauber J., Lazarevic M.,
RA Lee S.M., Levine A., Liu H., Masuda S., Mauel C., Medigue C.,
RA Medina N., Mellado R.P., Mizuno M., Moestl D., Nakai S., Noback M.,
RA Noone D., O'Reilly M., Ogawa K., Ogiwara A., Oudega B., Park S.H.,
RA Parro V., Pohl T.M., Portetelle D., Porwollik S., Prescott A.M.,
RA Presecan E., Pujic P., Purnelle B., Rapoport G., Rey M., Reynolds S.,
RA Rieger M., Rivolta C., Rocha E., Roche B., Rose M., Sadaie Y.,
RA Sato T., Scanlan E., Schleich S., Schroeter R., Scoffone F.,
RA Sekiguchi J., Sekowska A., Serror S.J., Serror P., Shin B.S., Solido B.,
RA Sorokin A., Tacconi E., Takagi T., Takahashi H., Takemaru K.,
```

RA Takeuchi M., Takakoshi A., Tanaka T., Terpstra P., Tognoni A.,
RA Tosato V., Uchiyama S., Vandenbol M., Vannier F., Vassarotti A.,
RA Viri A., Wambui R., Wedler E., Wedler H., Weitzenecker T.,
RA Winters P., Wipat A., Yamamoto H., Yamane K., Yasumoto K., Yata K.,
RA Yoshida K., Yoshikawa H.F., Zumbstein E., Yoshikawa H., Danchin A.,
RT "The complete genome sequence of the Gram-positive bacterium *Bacillus*
RT *subtilis*.";
RL Nature 390:249-256(1997).
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; L47709; AAB38433.1; -;
DR EMBL; Z99115; CAB14174.1; -;
DR Subtilist; BG11497; YpIB.
KW Hypothetical protein; Complete proteome.
SQ SEQUENCE 179 AA; 21364 MW; 46366E994DB709D9 CRC64;

Query Match 1.4%; Score 7; DB 1; Length 179;
Best Local Similarity 100.0%; Pred. No. 29;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 46 EQILEHA 52
DB 137 EQILEHA 143
|||||

RESULT 13
WCAF_ECOLI
ID WCAF_ECOLI STANDARD; PRT; 182 AA.
AC P71240; P76383;
DT 01-NOV-1997 (Rel. 35, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Putative colanic acid biosynthesis acetyltransferase wcaf
DE (EC 2.3.1.-).
GN WCAF OR B2054.
OS Escherichia coli.
OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
OC Escherichia.
OX NCBI_TaxID=562;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-K12;
RX MEDLINE=96326333; PubMed=8759852;
RA Stevenson G., Andrianopoulos K., Hobbs M., Reeves P.R.;
RT "Organization of the *Escherichia coli* K-12 gene cluster responsible
RT for production of the extracellular polysaccharide colanic acid.";
RL J. Bacteriol. 178:4885-4893(1996).
RN [2]
RP REVISIONS TO N-TERMINUS.
RC STRAIN-K12;
RA Reeves P.R.;
RL Submitted (APR-1998) to the EMBL/GenBank/DBJ databases.
RN [3]
RP SEQUENCE FROM N.A.
RX STRAIN-K12 / MG1655;
RX MEDLINE=97426617; PubMed=9278503;
RA Blattner F.R., Plunkett G. III, Bloch C.A., Perna N.T., Burland V.,
RA Riley M., Collado-Vides J., Glasner J.D., Rode C.K., Mayhew G.F.,
RA Gregor J., Davis N.W., Kirkpatrick H.A., Goeden M.A., Rose D.J.,
RA Mau B., Shao Y.;
RT "The complete genome sequence of *Escherichia coli* K-12.";
RL Science 277:1453-1474(1997).
RN [4]
RP SEQUENCE FROM N.A.
RC STRAIN-K12;
RX MEDLINE=97251358; PubMed=9097040;

RA Itoh T., Alba H., Baba T., Fujita K., Hayashi K., Inada T.,
RA Isono K., Kasai H., Kimura S., Kitakawa M., Kitagawa M.,
RA Makino K., Miki T., Mizobuchi K., Mori H., Mori T., Motomura K.,
RA Nakade S., Nakamura Y., Nashimoto H., Nishio Y., Oshima T.,
RA Saito N., Sempel G., Seki Y., Sivasundaram S., Tagami H.,
RA Takeda J., Takemoto K., Wada C., Yamamoto Y., Horiuchi T.;
RT "A 460-kb DNA sequence of the *Escherichia coli* K-12 genome
RT corresponding to the 40.1-50.0 min region on the linkage map.";
RL DNA Res. 3:379-392(1996).
CC -----
CC -!- PATHWAY: INVOLVED IN THE BIOSYNTHESIS OF THE SLIME POLYSACCHARIDE
CC COLANIC ACID.
CC -----
CC -!- SIMILARITY: BELONGS TO THE CYSE/LACA/LPXA/NODL FAMILY OF
CC ACETYLTRANSFERASES. COMPOSED OF MULTIPLE REPEATS OF [LIV]-G-X(4).
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; U38473; AAC77841.1; -;
DR EMBL; AE000295; AAC75115.1; -;
DR EMBL; D90843; BAA15910.1; -;
DR EcoGene; EGI3574; wcaf.
DR InterPro; IPR001451; Hexapep_transf.
DR Pfam; PF00132; hexapep; 3.
DR PROSITE; PS00101; HEXAPEP_TRANSFERASES; 1.
KW Lipopolysaccharide biosynthesis; Transferase; Acyltransferase; Repeat;
KW Complete proteome.
FT CONFLICT 50 F -> L (IN REF. 2).
SQ SEQUENCE 182 AA; 19962 MW; 8AA5E672CA802442 CRC64;

Query Match 1.4%; Score 7; DB 1; Length 182;
Best Local Similarity 100.0%; Pred. No. 29;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 8 KSLPANV 14
DB 159 KSLPANV 165
|||||

RESULT 14
DGC6_CHICK
ID DGC6_CHICK STANDARD; PRT; 200 AA.
AC O73770;
DT 30-MAY-2000 (Rel. 39, Created)
DT 30-MAY-2000 (Rel. 39, Last sequence update)
DT 30-MAY-2000 (Rel. 39, Last annotation update)
DE DGC6 protein (Digeorge syndrome critical region 6 homolog).
GN DGC6.
OS Gallus gallus (Chicken).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Archosauria; Aves; Neognathae; Galliformes; Phasianidae; Phasianinae;
OC Gallus.
OX NCBI_TaxID=9031;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-Broiler; TISSUE=Heart;
RA Hierck B.P., Poelmann R.E., Gittenberger-De Groot A.C.;
RL Submitted (FEB-1998) to the EMBL/GenBank/DBJ databases.
CC -!- SIMILARITY: BELONGS TO THE GONADAL FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; AF048985; AAC05142.1; -;

KW Coiled coil. 116 146 COILED COIL (POTENTIAL).
FT DOMAIN 176 198 COILED COIL (POTENTIAL).
SQ SEQUENCE 200 AA; 23242 MW; 7037BFF6F660E469 CRC64;

Query Match 1.4%; Score 7; DB 1; Length 200;
Best Local Similarity 100.0%; Pred. No. 32;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 462 ELDQKVI 468
Db 146 ELDQKVI 152

RESULT 15

GTTL_DROER STANDARD; PRT; 200 AA.
AC P30104;
DT 01-APR-1993 (Rel. 25, Created)
DT 01-APR-1993 (Rel. 25, Last sequence update)
DT 01-FEB-1996 (Rel. 33, Last annotation update)
DE Glutathione S-transferase 1-1 (EC 2.5.1.18) (GST class-theta)
DE (Fragment).
GN GSTD1 OR GST1 OR GST.
OS Drosophila erecta (Fruit fly).
OC Eukaryota; Metazoa; Arthropoda; Mandibulata; Pancrustacea; Hexapoda;
OC Insecta; Pterygota; Neoptera; Endopterygota; Diptera; Brachycera;
OC Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.
OX NCBI_TaxID=7220;
RN [1]
RP SEQUENCE FROM N.A.
RA Hargis M.T., Cochrane B.J.;
RL Submitted (XXX-1992) to the EMBL/GenBank/DBJ databases.
CC -!- FUNCTION: CONJUGATION OF REDUCED GLUTATHIONE TO A WIDE NUMBER
CC OF EXOGENOUS AND ENDOGENOUS HYDROPHOBIC ELECTROPHILES.
CC -!- CATALYTIC ACTIVITY: RX + glutathione -> HX + R-S-glutathione.
CC -!- SUBUNIT: HOMODIMER.
CC -!- SIMILARITY: BELONGS TO THE GST SUPERFAMILY. THETA FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; M84576; -; NOT_ANNOTATED_CDS.
DR HSP; P30712; 1LJR.
DR FlyBase; FBgn0012266; Dere\Gstd1.
DR InterPro; IPR004046; GST_Cterm.
DR InterPro; IPR004045; GST_Nterm.
DR Pfam; PF00043; GST_C; 1.
DR Pfam; PF02798; GST_N; 1.
KW Transferase; Multigene family.
FT NON_TER . 1
SQ SEQUENCE 200 AA; 22680 MW; 82052B2068FA4895 CRC64;

Query Match 1.4%; Score 7; DB 1; Length 200;
Best Local Similarity 100.0%; Pred. No. 32;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 107 APADPEA 113
Db 113 APADPEA 119

Search completed: May 8, 2003, 16:42:56
Job time : 17 secs

THIS PAGE BLANK (USPTO)

GenCore version 5.1.4_p5_4578
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: May 8, 2003, 16:40:54 ; Search time 89 Seconds
(without alignments)
1122.841 Million cell updates/sec

Title: US-09-926-163B-2
Perfect score: 485
Sequence: 1 MITRETLKSLPANVQAPPYD.....KVIVLRKIIRKGVKAAIPA 485

Scoring table: OLIGO
Gapop 60.0 , Gapext 60.0

Searched: 671580 seqs, 206047115 residues

Word size : 0
Total number of hits satisfying chosen parameters: 671580

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Database : SPTREMBL_21.*

- 1: sp_archaea.*
- 2: sp_bacteria.*
- 3: sp_fungi.*
- 4: sp_human.*
- 5: sp_invertebrate.*
- 6: sp_mammal.*
- 7: sp_mhc.*
- 8: sp_organelle.*
- 9: sp_phase.*
- 10: sp_plant.*
- 11: sp_rodent.*
- 12: sp_virus.*
- 13: sp_vertebrate.*
- 14: sp_unclassified.*
- 15: sp_ivirus.*
- 16: sp_bacteriap.*
- 17: sp_archaeap.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	485	100.0	485	2 Q9KWR5	Q9kwr5 gluconobact
2	52	10.7	485	2 Q8VUU5	Q8vuus5 gluconobact
3	13	2.7	525	16 Q8U7J8	Q8u7j8 agrobacteri
4	11	2.3	492	16 Q98D06	Q98d06 rhizobium l
5	11	2.3	494	16 Q92N07	Q92n07 rhizobium m
6	10	2.1	486	16 Q8X516	Q8x516 escherichia
7	10	2.1	491	16 Q911D6	Q91ld6 pseudomonas
8	10	2.1	493	2 Q8B355	Q8b355 pseudomonas
9	9	1.9	486	16 Q8X506	Q8x506 escherichia
10	9	1.9	488	16 Q8ZP18	Q8zpi8 salmonella
11	9	1.9	488	16 Q8Z626	Q8z626 salmonella
12	8	1.6	244	16 Q8UD98	Q8ud98 agrobacteri
13	8	1.6	253	16 P74566	P74566 synecocyst
14	8	1.6	261	5 Q9W439	Q9w439 drosophila
15	8	1.6	264	2 Q24792	Q24792 borrelia ga
16	8	1.6	301	16 Q8XUP2	Q8xup2 ralstonia s

17	1.6	374	16	Q9L1S8	Q9l1s8 streptomyce
18	1.6	409	16	Q9RJ07	Q9rj07 streptomyce
19	1.6	580	16	Q9CF08	Q9cf08 lactococcus
20	1.6	1006	10	Q9LT95	Q9lt95 arabidopsis
21	1.6	1030	3	Q9HFQ9	Q9hfg9 emericella
22	1.4	18	6	P79214	P79214 cryptotlagus
23	1.4	42	9	Q9AYV1	Q9ayv1 lactococcus
24	1.4	42	9	Q38332	Q38332 lactococcus
25	1.4	42	9	Q9B011	Q9b011 bacterioph
26	1.4	55	16	Q8XYZ1	Q8xyz1 ralstonia s
27	1.4	72	16	Q9JR22	Q9jr22 neisseria m
28	1.4	106	10	Q94IU3	Q94iu3 fragaria nu
29	1.4	106	10	Q94IU3	Q94iu3 fragaria ve
30	1.4	110	17	Q9HKU9	Q9hku9 thermoplasm
31	1.4	128	10	Q93X92	Q93x92 lotus corni
32	1.4	130	16	Q8YZC5	Q8yzc5 anabaena sp
33	1.4	139	16	Q92RA2	Q92ra2 rhizobium m
34	1.4	142	16	P73486	P73486 synecocyst
35	1.4	142	16	Q9AC17	Q9ac17 caulobacter
36	1.4	147	17	Q976M0	Q976m0 sulfolobus
37	1.4	148	17	Q8U067	Q8u067 pyrococcus
38	1.4	149	9	O80313	O80313 bacterioph
39	1.4	149	16	Q98RB3	Q98rb3 mycoplasma
40	1.4	154	16	O06744	O06744 bacillus su
41	1.4	158	4	Q96D43	Q96d43 homo sapien
42	1.4	158	5	Q9NMK5	Q9nmk5 leishmania
43	1.4	164	16	Q98GU0	Q98gu0 rhizobium l
44	1.4	165	10	Q946W6	Q946w6 prunus pers
45	1.4	166	16	Q9KA62	Q9ka62 bacillus ha

ALIGNMENTS

RESULT 1

Q9KWR5 ID Q9KWR5 PRELIMINARY; PRT; 485 AA.
AC Q9KWR5;
DT 01-OCT-2000 (TREMREL. 15, Created)
DT 01-OCT-2000 (TREMREL. 15, Last sequence update)
DT 01-OCT-2001 (TREMREL. 18, Last annotation update)
DE Sorbitol dehydrogenase.
OS Gluconobacter oxydans (Gluconobacter suboxydans).
OC Bacteria; Proteobacteria; alpha subdivision; Acetobacteraceae;
OC Gluconobacter.
OX NCBI_TaxID=442;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=G624;
RA Shibata T., Ichikawa C., Matsuura M., Takata Y., Noguchi Y., Saito Y., Yamashita M.;
RT "Cloning of a gene for D-Sorbitol dehydrogenase from Gluconobacter oxydans G624 and expression of the gene in Pseudomonas putida IF03738.";
RT J. Biosci. Bioeng. 89:463-468(2000).
RL EMBL; AB028937; BAA99414.1; .
DR InterPro: IPR000669; Mannitol_dh.
DR Pfam: PF01232; Mannitol_dh; 1.
DR PRINTS: PR00084; MTLDRGNASE.
SQ SEQUENCE 485 AA; 53642 MW; A7C363D728EEA7DC CRC64;

Query Match 100.0%; Score 485; DB 2; Length 485;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 485; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MITRETLKSLPANVQAPPYDIDGIRKPIVHFGVGNFFRAHEAFYVEQILEHAPDWAIVGV 60
|||||
Db 1 MITRETLKSLPANVQAPPYDIDGIRKPIVHFGVGNFFRAHEAFYVEQILEHAPDWAIVGV 60
|||||
QY 61 GLUTGSDRSKKAEFEKQAQCLYSLTETAPSGKSTVVMGALRDYLLAPADPAVLKHLVD 120
|||||
Db 61 GLUTGSDRSKKAEFEKQAQCLYSLTETAPSGKSTVVMGALRDYLLAPADPAVLKHLVD 120
|||||

```
QY 121 PAIRIVSMITTEGGYNINETTGAFDLENAAVKADLNKPEKSTVFYVVEALRRRDAGG 180
Db 121 PAIRIVSMITTEGGYNINETTGAFDLENAAVKADLNKPEKSTVFYVVEALRRRDAGG 180
QY 181 KAFTVMSCDNLRHNGNVARKAFLGYAKARDEPDLAKWIEENATFPNGWDRITPTVSAETA 240
Db 181 KAFTVMSCDNLRHNGNVARKAFLGYAKARDEPDLAKWIEENATFPNGWDRITPTVSAETA 240
QY 241 KKLNAASGIDDDLLPLVAEDFHOWVLEDOFADGRPPLEKAGVOMGVDVTDWYVKIRMLNA 300
Db 241 KKLNAASGIDDDLLPLVAEDFHOWVLEDOFADGRPPLEKAGVOMGVDVTDWYVKIRMLNA 300
QY 301 GHVMLCFPGILGYENVDDAIEDSELLGNLKNYLNKDVIPTLKAPSGMTLEGYRDSVISR 360
Db 301 GHVMLCFPGILGYENVDDAIEDSELLGNLKNYLNKDVIPTLKAPSGMTLEGYRDSVISR 360
QY 361 FSNKMSDQTLRIASDCGSKVQVFWTETVYRRATIEDKRLSRIAFGSIASYLEMLGRDEKG 420
Db 361 FSNKMSDQTLRIASDCGSKVQVFWTETVYRRATIEDKRLSRIAFGSIASYLEMLGRDEKG 420
QY 421 GTYESSEPTYGDAEWKLAKADDFESSKLPLAFDGNRDLDTSELDOKVIVLRIIREKGVK 480
Db 421 GTYESSEPTYGDAEWKLAKADDFESSKLPLAFDGNRDLDTSELDOKVIVLRIIREKGVK 480
QY 481 AAIPA 485
Db 481 AAIPA 485

RESULT 2
Q8VU05
ID Q8VU05 PRELIMINARY; PRT: 485 AA.
AC Q8VU05;
DT 01-WAR-2002 (TREMBLrel. 20, Created)
DT 01-WAR-2002 (TREMBLrel. 20, Last sequence update)
DT 01-JUN-2002 (TREMBLrel. 21, Last annotation update)
DE L-sorbose reductase.
OS Gluconobacter oxydans (Gluconobacter suboxydans).
OC Bacteria; Proteobacteria; alpha subdivision; Acetobacteraceae;
OX Gluconobacter.
OX NCBI_TaxID=442;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=IFO3291;
RX MEDLINE=21650687; PubMed=11790761;
RA Shinjoh M., Tazoe M., Hoshino T.;
RT "NADPH-Dependent L-Sorbose Reductase Is Responsible for L-Sorbose
RT Assimilation in Gluconobacter suboxydans IFO 3291.";
RL J. Bacteriol. 184:861-863(2002).
DR EMBL; AB063188; BAB83933.1; -.
DR InterPro; IPR000669; Mannitol_dh.
DR Pfam; PF01232; Mannitol_dh; 1.
DR PRINTS; PR00084; MTLDDHGRNASE.
SQ SEQUENCE 485 AA; 53541 MW; 883EC4956B3C6CF0 CRC64;

Query Match 10.7%; Score 52; DB 2; Length 485;
Best Local Similarity 100.0%; Pred. No. 1.4e-44;
Matches 52; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 186 MSCDNLRHNGNVARKAFLGYAKARDEPDLAKWIEENATFPNGWDRITPTVSA 237
Db 186 MSCDNLRHNGNVARKAFLGYAKARDEPDLAKWIEENATFPNGWDRITPTVSA 237

RESULT 3
Q8U7J8
ID Q8U7J8 PRELIMINARY; PRT: 525 AA.
AC Q8U7J8;
DT 01-JUN-2002 (TREMBLrel. 21, Created)
DT 01-JUN-2002 (TREMBLrel. 21, Last sequence update)
DT 01-JUN-2002 (TREMBLrel. 21, Last annotation update)
DE Mannitol 2-dehydrogenase.
```

```
GN MTLK OR ATU4451 OR AGR_L_830.
OS Agrobacterium tumefaciens (strain C58 / ATCC 33970).
OC Bacteria; Proteobacteria; alpha subdivision; Rhizobiaceae group;
OC Rhizobiaceae; Rhizobium.
OX NCBI_TaxID=176299;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=21608550; PubMed=11743193;
RA Wood D.W., Setubal J.C., Kaul R., Monks D.E., Kitajima J.P.,
RA Okura Y.K., Zhou Y., Chen L., Wood G.E., Almeida N.F. Jr., Woo L.,
RA Chen Y., Paulsen I.T., Eisen J.A., Karp P.D., Bovee D. Sr.,
RA Chapman P., Clendenning J., Deatherage G., Gillet W., Grant C.,
RA Kutyavlin T., Levy R., Li M.-J., McClelland E., Palmieri A.,
RA Raymond C., Rouse G., Saenphimmachak C., Wu Z., Romero P., Gordon D.,
RA Zhang S., Yoo H., Tao Y., Biddle P., Jung M., Krespan W., Perry M.,
RA Gordon-Kamm B., Liao L., Kim S., Hendrick C., Zhao Z.-Y., Dolan M.,
RA Chumley F., Tingey S.V., Tomb J.-F., Gordon M.P., Olson M.V.,
RA Nestor E.W.;
RT "The genome of the natural genetic engineer Agrobacterium tumefaciens
RT C58.";
RN [2]
RP Science 294:2317-2323(2001).
RN [2]
RP SEQUENCE FROM N.A.
RX MEDLINE=21608551; PubMed=11743194;
RA Goodner B., Hinkle G., Gattung S., Miller N., Blanchard M.,
RA Qureshi B., Goldman B.S., Cao Y., Askenazi M., Halling C., Mullin L.,
RA Houmlel K., Gordon J., Vaudin M., Iartchouk O., Epp A., Liu F.,
RA Wollam C., Allinger M., Dougherty D., Scott C., Lappas C., Markelz B.,
RA Flanagan C., Crowell C., Gurson J., Lomo C., Sear C., Strub G.,
RA Cielo C., Slater S.;
RT "Genome sequence of the plant pathogen and biotechnology agent
RT Agrobacterium tumefaciens C58.";
RN [2]
RP Science 294:2323-2328(2001).
DR EMBL; AE009374; AAL45245.1; ALT_INIT.
DR EMBL; AE008240; AAK88989.1; -.
KW Complete proteome.
SQ SEQUENCE 525 AA; 57859 MW; F65432D35A4AFA6C CRC64;

Query Match 2.7%; Score 13; DB 16; Length 525;
Best Local Similarity 100.0%; Pred. No. 0.00027;
Matches 13; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 123 IRIVSMITTEGGY 135
Db 155 IRIVSMITTEGGY 167

RESULT 4
Q98D06
ID Q98D06 PRELIMINARY; PRT: 492 AA.
AC Q98D06;
DT 01-OCT-2001 (TREMBLrel. 18, Created)
DT 01-OCT-2001 (TREMBLrel. 18, Last sequence update)
DT 01-JUN-2002 (TREMBLrel. 21, Last annotation update)
DE Mannitol dehydrogenase.
GN MLL4920.
OS Rhizobium loti (Mesorhizobium loti).
OC Bacteria; Proteobacteria; alpha subdivision; Rhizobiaceae group;
OC Phyllobacteriaceae; Mesorhizobium.
OX NCBI_TaxID=381;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=21082930; PubMed=11214968;
RA Kaneko T., Nakamura Y., Sato S., Asamizu E., Kato T., Sasamoto S.,
RA Watanabe A., Idesawa K., Ishikawa A., Kawashima K., Kimura T.,
RA Kishida Y., Kiyokawa C., Kohara M., Matsumoto M., Matsuno A.,
RA Mochizuki Y., Nakayama S., Nakazaki N., Shimpo S., Sugimoto M.,
RA Takeuchi C., Yamada M., Tabata S.;
RT "Complete genome structure of the nitrogen-fixing symbiotic bacterium
RT Mesorhizobium loti.";
RN [1]
RP DNA Res. 7:331-338(2000).
DR EMBL; AP003005; BAB51465.1; -.
```

```

DR InterPro; IPR000669; Mannitol_dh.
DR Pfam; PF01232; Mannitol_dh; 1.
DR PRINTS; PR00084; MTLDRGNASE.
DR PROSITE; PS00974; MANNITOL_DHGENASE; 1.
KW COMPLETE PROTEOME.
SQ SEQUENCE 492 AA; 53518 MW; 4D20E1136C3C1373 CRC64;

Query Match          2.3%; Score 11; DB 16; Length 492;
Best Local Similarity 100.0%; Pred. No. 0.03;
Matches 11; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 223 FPNQWVDRIPT 233
DB 224 FPNQWVDRIPT 234

RESULT 5
Q92N07 PRELIMINARY; PRT; 494 AA.
AC Q92N07;
DT 01-DEC-2001 (TREMBLrel. 19, Created)
DT 01-DEC-2001 (TREMBLrel. 19, Last sequence update)
DT 01-MAR-2002 (TREMBLrel. 20, Last annotation update)
DE Probable mannitol 2-dehydrogenase protein (EC 1.1.1.67).
GN MTLK OR R02440 OR SMC01501.
OS Rhizobium meliloti (Sinorhizobium meliloti).
OC Bacteria; Proteobacteria; alpha subdivision; Rhizobiaceae group;
OC Rhizobiaceae; Sinorhizobium.
OX NCBI_TaxID=382;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=1021;
RX MEDLINE=21396507; PubMed=11481430;
RA Capela D., Barloy-Hubler F., Gouzy J., Bothe G., Ampe F., Batut J.,
RA Boistard P., Becker A., Boutry M., Cadieu E., Dreano S., Gloux S.,
RA Gouzie T., Goffeau A., Kahn D., Kiss E., Lelaure V., Masuy D.,
RA Pohl T., Portetelle D., Puchler A., Purnelle B., Rampsperger U.,
RA Renard C., Thebaud P., Vandebol M., Weidner S., Galibert F.;
RT "Analysis of the chromosome sequence of the legume symbiont
RT Sinorhizobium meliloti strain 1021.";
RL Proc. Natl. Acad. Sci. U.S.A. 98:9877-9882(2001).
DR EMBL; AL591790; CAC47019.1;
DR InterPro; IPR000669; Mannitol_dh.
DR Pfam; PF01232; Mannitol_dh; 1.
DR PROSITE; PS00974; MANNITOL_DHGENASE; UNKNOWN_1.
KW Oxidoreductase; Complete proteome.
SQ SEQUENCE 494 AA; 53822 MW; 7533287D8EF58F24 CRC64;

Query Match          2.3%; Score 11; DB 16; Length 494;
Best Local Similarity 100.0%; Pred. No. 0.03;
Matches 11; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 223 FPNQWVDRIPT 233
DB 224 FPNQWVDRIPT 234

RESULT 6
Q8X516 PRELIMINARY; PRT; 486 AA.
AC Q8X516;
DT 01-MAR-2002 (TREMBLrel. 20, Created)
DT 01-MAR-2002 (TREMBLrel. 20, Last sequence update)
DT 01-JUN-2002 (TREMBLrel. 21, Last annotation update)
DE Mannonate oxidoreductase (D-mannonate oxidoreductase).
GN UXUB OR Z5921 OR EC55282.
OS Escherichia coli O157:H7.
OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
OC Escherichia.
OX NCBI_TaxID=83334;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=O157:H7 / EDL933 / ATCC 700927;

```

```

RX MEDLINE=21074935; PubMed=11206551;
RA Perna N.T., Plunkett G. III, Burland V., Mau B., Glasner J.D.,
RA Rose D.J., Mayhew G.F., Evans P.S., Gregor J., Kirkpatrick H.A.,
RA Posfai G., Hackett J., Klink S., Boutin A., Shao Y., Miller L.,
RA Grotbeck E.J., Davis N.W., Lim A., Dimalanta E.T., Potamoudis K.,
RA Apodaca J., Anantharaman T.S., Lin J., Yen G., Schwartz D.C.,
RA Welch R.A., Blattner F.R.;
RT "Genome sequence of enterohaemorrhagic Escherichia coli O157:H7.";
RL Nature 409:529-533(2001).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=O157:H7 / RIMD 0509952;
RX MEDLINE=21156231; PubMed=11258796;
RA Hayashi T., Makino K., Ohnishi M., Kurokawa K., Ishii K., Yokoyama K.,
RA Han C.-G., Ohtsubo E., Nakayama K., Murata T., Tanaka M., Tobe T.,
RA Iida T., Takami H., Honda T., Sasakawa C., Ogasawara N., Yasunaga T.,
RA Kuhara S., Shiba T., Hattori M., Shinagawa H.;
RT "Complete genome sequence of enterohaemorrhagic Escherichia coli
RT O157:H7 and genomic comparison with a laboratory strain K-12.";
RL DNA Res. 8:11-22(2001).
DR EMBL; AE005663; AAG59505.1;
DR EMBL; AP002569; BAB38705.1;
DR InterPro; IPR000669; Mannitol_dh.
DR Pfam; PF01232; Mannitol_dh; 1.
DR PRINTS; PR00084; MTLDRGNASE.
DR PROSITE; PS00974; MANNITOL_DHGENASE; 1.
KW Complete proteome.
SQ SEQUENCE 486 AA; 53656 MW; C519DD28763A41DC CRC64;

Query Match          2.1%; Score 10; DB 16; Length 486;
Best Local Similarity 100.0%; Pred. No. 0.32;
Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 181 KAFTVNSCDN 190
DB 178 KAFTVNSCDN 187

RESULT 7
Q91ID6 PRELIMINARY; PRT; 491 AA.
AC Q91ID6;
DT 01-MAR-2001 (TREMBLrel. 16, Created)
DT 01-MAR-2001 (TREMBLrel. 16, Last sequence update)
DT 01-MAR-2002 (TREMBLrel. 20, Last annotation update)
DE Mannitol dehydrogenase.
GN MTLDR PA2342.
OS Pseudomonas aeruginosa.
OC Bacteria; Proteobacteria; gamma subdivision; Pseudomonadaceae;
OC Pseudomonas.
OX NCBI_TaxID=287;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=ATCC 15692 / PA01;
RX MEDLINE=20437337; PubMed=10984043;
RA Stover C.K., Pham X.-Q.T., Erwin A.L., Mizoguchi S.D., Warren P.,
RA Hickey M.J., Brinkman F.S.L., Hufnagle W.O., Kowalik D.J., Lagrou M.,
RA Garber R.L., Goltry L., Tolentino E., Westbrook-Wadman S., Lim Y.,
RA Brody L.L., Coulter S.N., Folger K.R., Kas A., Larbig K., Lim R.M.,
RA Smith K.A., Spencer D.H., Wong G.K.-S., Wu Z., Paulsen I.T.,
RA Reizer J., Sailer M.H., Hancock R.E.W., Lory S., Olson M.V.;
RT "Complete genome sequence of Pseudomonas aeruginosa PA01, an
RT opportunistic pathogen.";
RL Nature 406:959-964(2000).
DR EMBL; AE004660; AAG05730.1;
DR InterPro; IPR000669; Mannitol_dh.
DR Pfam; PF01232; Mannitol_dh; 1.
DR PRINTS; PR00084; MTLDRGNASE.
DR PROSITE; PS00974; MANNITOL_DHGENASE; 1.
KW Complete proteome.
SQ SEQUENCE 491 AA; 54350 MW; 8729A9F7E9841C39 CRC64;

Query Match          2.1%; Score 10; DB 16; Length 491;

```

```

Best Local Similarity 100.0%; Pred. No. 0.32; Mismatches 0; Indels 0; Gaps 0;
Matches 10; Conservative 0;

Qy 182 AFTVMSCDNL 191
Db 183 AFTVMSCDNL 192
|||||

RESULT 8
O08355 PRELIMINARY; PRT; 493 AA.
AC O08355;
DT 01-JUL-1997 (TReMBLrel. 04, Created)
DT 01-JUL-1997 (TReMBLrel. 04, Last sequence update)
DT 01-JUN-2001 (TReMBLrel. 17, Last annotation update)
DE Mannitol 2-dehydrogenase (EC 1.1.1.67) (MDH).
CN MTD.
OS Pseudomonas fluorescens.
OC Bacteria; Proteobacteria; gamma subdivision; Pseudomonadaceae;
OC Pseudomonas.
OX NCBI_TaxID=294;
RN [1]
RP SEQUENCE FROM N.A., CHARACTERIZATION, AND SEQUENCE OF 1-20.
RC STRAIN=DSM 50106;
RX MEDLINE=97236441; PubMed=9116029;
RA Bruenker P., Altenbuchner J., Kulbe K.D., Mattes R.;
RT "Cloning, nucleotide sequence and expression of a mannitol
dehydrogenase gene from Pseudomonas fluorescens DSM 50106 in
Escherichia coli.";
RT Escherichia coli.;
RL Biochim. Biophys. Acta 1351:157-167(1997).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=DSM 50106;
RX MEDLINE=98121321; PubMed=9461423;
RA Bruenker P., Altenbuchner J., Mattes R.;
RT "Structure and function of the genes involved in mannitol, arabinol
and glucitol utilization from Pseudomonas fluorescens DSM50106.";
RL Gene 206:117-126(1998).
CC -!- FUNCTION: A MANNITOL DEHYDROGENASE WITH A BROAD SUBSTRATE
SPECIFICITY. SUBSTRATES INCLUDE MANNITOL, ARABITOL AND SORBITOL,
THESE ARE OXIDIZED TO GIVE THE CORRESPONDING KETO SUGARS. THIS
ENZYME WILL ALSO CATALYZE THE REDUCTION OF FRUCTOSE AND XYLULOSE.
CC -!- CATALYTIC ACTIVITY: D-MANNITOL + NAD(+) -> D-FRUCTOSE + NADH.
CC -!- SUBUNIT: MONOMER.
CC -!- INDUCTION: BY MANNITOL, ARABITOL AND SORBITOL. NOT INDUCED BY
FRUCTOSE OR XYLULOSE.
CC -!- MISCELLANEOUS: THE MTDL PROTEIN IS ENCODED BY THE MTL EFGKDYZ
OPERON. THIS OPERON ENCODES PROTEINS FOR THE UPTAKE AND
UTILIZATION OF MANNITOL, ARABITOL AND SORBITOL.
CC -!- SIMILARITY: BELONGS TO THE MANNITOL DEHYDROGENASES FAMILY.
DR EMBL; AF007800; AAC04472.1; -.
DR InterPro; IPR000669; Mannitol_dh.
DR Pfam; PF01232; Mannitol_dh; 1.
DR PRINTS; PR00084; MTDLHDHGNASE.
DR PROSITE; PS00974; MANNITOL_DHGENASE; 1.
KW Oxidoreductase; NAD.
FT NP_BIND 29 40
SQ SEQUENCE 493 AA; 54497 MW; 7C12DFA443CEA443 CRC64;

Query Match 2.1%; Score 10; DB 2; Length 493;
Best Local Similarity 100.0%; Pred. No. 0.32;
Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 182 AFTVMSCDNL 191
Db 183 AFTVMSCDNL 192
|||||

RESULT 9
Q8X506 PRELIMINARY; PRT; 486 AA.
AC Q8X506;
DT 01-MAR-2002 (TReMBLrel. 20, Created)

```

```

DT 01-MAR-2002 (TReMBLrel. 20, Last sequence update)
DT 01-JUN-2002 (TReMBLrel. 21, Last annotation update)
DE Putative oxidoreductase.
GN YDFI OR Z2155 OR ECS2151.
OS Escherichia coli O157:H7.
OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
OC Escherichia.
OX NCBI_TaxID=83334;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=O157:H7 / EDL933 / ATCC 700927;
RX MEDLINE=21074935; PubMed=11206551;
RA Perna N.T., Plunkett G. III, Burland V., Mau B., Glasner J.D.,
Rose D.J., Mayhew G.F., Evans P.S., Gregor J., Kirkpatrick H.A.,
Posafai G., Hackett J., Klink S., Boutin A., Shao Y., Miller L.,
Apodaca E.J., Davis N.W., Lim A., Dimalanta E.T., Potamoudis K.,
Apodaca J., Anantharaman T.S., Lin J., Yen G., Schwartz D.C.,
Welch R.A., Blattner F.R.;
RT "Genome sequence of enterohaemorrhagic Escherichia coli O157:H7.";
RL Nature 409:529-533(2001).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=O157:H7 / RIMD 050952;
RX MEDLINE=21156231; PubMed=11258796;
RA Hayashi T., Makino K., Ohnishi M., Kurokawa K., Ishii K., Yokoyama K.,
Han C.-G., Ohtsubo E., Nakayama K., Murata T., Tanaka M., Tobe T.,
Iida T., Takami H., Honda T., Sasaki K., Ogasawara N., Yasunaga T.,
Kuhara S., Shiba T., Hattori M., Shinagawa H.;
RT "Complete genome sequence of enterohaemorrhagic Escherichia coli
O157:H7 and genomic comparison with a laboratory strain K-12.";
RL DNA Res. 8:11-22(2001).
DR EMBL; AE005350; AAG56220.1; -.
DR EMBL; AP002557; BAB35574.1; -.
DR InterPro; IPR000669; Mannitol_dh.
DR Pfam; PF01232; Mannitol_dh; 1.
DR PRINTS; PR00084; MTDLHDHGNASE.
DR PROSITE; PS00974; MANNITOL_DHGENASE; 1.
KW Complete proteome.
SQ SEQUENCE 486 AA; 53581 MW; 1AFD234D83108D9E CRC64;

Query Match 1.9%; Score 9; DB 16; Length 486;
Best Local Similarity 100.0%; Pred. No. 3.4;
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 182 AFTVMSCDNL 190
Db 176 AFTVMSCDNL 184
|||||

RESULT 10
Q8ZPI8 PRELIMINARY; PRT; 488 AA.
AC Q8ZPI8;
DT 01-MAR-2002 (TReMBLrel. 20, Created)
DT 01-MAR-2002 (TReMBLrel. 20, Last sequence update)
DT 01-JUN-2002 (TReMBLrel. 21, Last annotation update)
DE Putative mannitol dehydrogenase.
GN YDFI OR STM1508.
OS Salmonella typhimurium.
OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
OC Salmonella.
OX NCBI_TaxID=602;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=LT2 / SGSC1412 / ATCC 700720;
RX MEDLINE=21534948; PubMed=11677609;
RA McClelland M., Sanderson K.E., Spieth J., Clifton S.W., Latreille P.,
Courtney L., Porwollik S., Ali J., Dante M., Du F., Hou S., Layman D.,
Leonard S., Nguyen C., Scott K., Holmes A., Grewal N., Mulvaney E.,
Ryan E., Sun H., Florea L., Miller W., Stoneking T., Nhan M.,
Waterston R., Wilson R.K.;
RT "Complete genome sequence of Salmonella enterica serovar Typhimurium
LT2.";

```

```

RL Nature 413:852-856(2001).
DR EMBL: AE008765; AAL20427.1; -.
DR InterPro: IPR000669; Mannitol_dh.
DR Pfam: PF01232; Mannitol_dh; 1.
DR PRINTS: PR00084; MTLDHDRGNASE.
DR PROSITE: PS00974; MANNITOL_DHGENASE; 1.
KW Hypothetical protein; Complete proteome.
SQ SEQUENCE 488 AA; 53896 MW; 704BF8C27F4407A6 CRC64;

Query Match
Best Local Similarity 100.0%; Score 9; DB 16; Length 488;
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 182 AFTVMSCDN 190
Db 176 AFTVMSCDN 184

RESULT 11
Q82626 ID Q82626 PRELIMINARY; PRT; 488 AA.
AC Q82626;
DT 01-MAR-2002 (TrEMBLrel. 20, Created)
DT 01-MAR-2002 (TrEMBLrel. 20, Last sequence update)
DT 01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
DE Putative D-mannonate oxidoreductase (EC 1.1.1.57).
GN STV1553.
OS Salmonella typhi.
OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
OC Salmonella.
OX NCBI_TaxID=601;
RN STRAIN=CT18;
RP SEQUENCE FROM N.A.
RC MEDLINE=21534947; PubMed=11677608;
RA Parkhill J., Dougan G., James K.D., Thomson N.R., Pickard D., Wain J.,
RA Churcher C., Mungall K.L., Bentley S.D., Holden M.T.G., Sebaihia M.,
RA Baker S., Basham D., Brooks K., Chillingworth T., Connor P.,
RA Cronin A., Davis P., Davies R.M., Dowd L., White N., Farrar J.,
RA Feltwell T., Hamlin N., Haque A., Hien T.T., Holroyd S., Jagels K.,
RA Krogh A., Larsen T.S., Leather S., Moule S., O'Gaora P., Parry C.,
RA Quail M., Rutherford K., Simmonds M., Skelton J., Stevens K.,
RA Whitehead S., Barrall B.G.;
RT "Complete genome sequence of a multiple drug resistant Salmonella
RT enterica serovar Typhi CT18.";
RL Nature 413:848-852(2001).
DR EMBL: AL627270; CAD01805.1; -.
DR InterPro: IPR000669; Mannitol_dh.
DR Pfam: PF01232; Mannitol_dh; 1.
DR PRINTS: PR00084; MTLDHDRGNASE.
DR PROSITE: PS00974; MANNITOL_DHGENASE; 1.
KW Oxidoreductase; Hypothetical protein; Complete proteome.
SQ SEQUENCE 488 AA; 53922 MW; 18F64DE2ACCA7781 CRC64;

Query Match
Best Local Similarity 100.0%; Score 9; DB 16; Length 488;
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 182 AFTVMSCDN 190
Db 176 AFTVMSCDN 184

RESULT 12
Q8UD98 ID Q8UD98 PRELIMINARY; PRT; 244 AA.
AC Q8UD98;
DT 01-JUN-2002 (TrEMBLrel. 21, Created)
DT 01-JUN-2002 (TrEMBLrel. 21, Last sequence update)
DT 01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
DE ABC transporter, substrate binding protein.
GN ATU2229 OR AGR_C_4053.
OS Agrobacterium tumefaciens (strain C58 / ATCC 33970).

```

```

OC Bacteria; Proteobacteria; alpha subdivision; Rhizobiaceae group;
OC Rhizobiaceae; Rhizobium.
OX NCBI_TaxID=176299;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=21608550; PubMed=11743193;
RA Wood D.W., Setubal J.C., Kaul R., Monks D.E., Kitajima J.P.,
RA Okura V.K., Zhou Y., Chen L., Wood G.E., Almeida N.F. Jr., Woo L.,
RA Chen Y., Paulsen I.T., Eisen J.A., Karp P.D., Bovee D. Sr.,
RA Chapman P., Clendenning J., Deatherage G., Gillet W., Grant C.,
RA Kutyavin T., Levy R., Li M.-J., McClelland E., Palmieri A., Gordon D.,
RA Raymond C., Rouse G., Saenphimmachak C., Wu Z., Romero P., Gordon D.,
RA Zhang S., Yoo H., Tao Y., Biddle P., Jung M., Krespan W., Perry M.,
RA Gordon-Kamm B., Liao L., Kim S., Hendrick C., Zhao Z.-Y., Dolan M.,
RA Chumley F., Tingey S.V., Tomb J.-F., Gordon M.P., Olson M.V.,
RA Nester E.W.;
RT "The genome of the natural genetic engineer Agrobacterium tumefaciens
RT C58.";
RL Science 294:2317-2323(2001).
RN [2]
RP SEQUENCE FROM N.A.
RX MEDLINE=21608551; PubMed=11743194;
RA Goodner B., Hinkle G., Gattung S., Miller N., Blanchard M.,
RA Qurolo B., Goldman B.S., Cao Y., Askenazi M., Halling C., Mullin L.,
RA Humiel K., Gordon J., Vaudin M., Iartchouk O., Epp A., Liu F.,
RA Wollam C., Allinger M., Dougherty D., Scott C., Lappas C., Markelz B.,
RA Flanagan C., Crowell C., Gursion J., Lomo C., Sear C., Strub G.,
RA Cielo C., Slater S.;
RT "Genome sequence of the plant pathogen and biotechnology agent
RT Agrobacterium tumefaciens C58.";
RL Science 294:2323-2328(2001).
DR EMBL: AE009172; AAL43218.1; -.
DR EMBL: AE008138; AAK87971.1; -.
KW Complete proteome.
SQ SEQUENCE 244 AA; 27305 MW; 459F7E15EA3FF3B CRC64;

Query Match
Best Local Similarity 100.0%; Score 8; DB 16; Length 244;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 6 TLKSLPAN 13
Db 183 TLKSLPAN 190

RESULT 13
P74566 ID P74566 PRELIMINARY; PRT; 253 AA.
AC P74566;
DT 01-FEB-1997 (TrEMBLrel. 02, Created)
DT 01-FEB-1997 (TrEMBLrel. 02, Last sequence update)
DT 01-MAR-2002 (TrEMBLrel. 20, Last annotation update)
DE Hypothetical protein sir0654.
GN SLR0654.
OS Synecocystis sp. (strain PCC 6803).
OC Bacteria; Cyanobacteria; Chroococcales; Synecocystis.
OX NCBI_TaxID=1148;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=97061201; PubMed=8905231;
RA Kaneko T., Sato S., Kotani H., Tanaka A., Asamizu E., Nakamura Y.,
RA Miyajima N., Hirose M., Sugita M., Sasamoto S., Kimura T.,
RA Hosouchi T., Mitsuno A., Muraki A., Nakazaki N., Naruo K., Okumura S.,
RA Shimpo S., Takeuchi C., Wada T., Watanabe A., Yamada M., Yasuda M.,
RA Tabata S.;
RT "Sequence analysis of the genome of the unicellular cyanobacterium
RT Synecocystis sp. strain PCC6803. II. Sequence determination of the
RT entire genome and assignment of potential protein-coding regions.";
RL DNA Res. 3:109-136(1996).
DR EMBL: D90916; BAA18673.1; -.
KW Hypothetical protein; Complete proteome.
SQ SEQUENCE 253 AA; 28428 MW; F836F5042295F03B CRC64;

```

Query Match 1.6%; Score 8; DB 16; Length 253;
Best Local Similarity 100.0%; Pred. No. 20;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 338 VIPTLKAP 345
|||||
Db 2 VIPTLKAP 9

RESULT 14
Q9W439 PRELIMINARY; PRT; 261 AA.
AC Q9W439;
DT 01-MAY-2000 (TRENBLrel. 13, Created)
DT 01-MAR-2001 (TRENBLrel. 16, Last sequence update)
DT 01-MAR-2002 (TRENBLrel. 20, Last annotation update)
DE CG4660 protein.
GN CG4660.
OS Drosophila melanogaster (Fruit fly).
OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
OC Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC Ephydroidea; Drosophilidae; Drosophila.
OX NCBI_TaxID=7227;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=BERKELEY;
RX MEDLINE=20196006; PubMed=10731132;
RA Adams M.D., Celniker S.E., Holt R.A., Evans C.A., Gocayne J.D.,
RA Amanatides P.G., Scher S.E., Li P.W., Hoskins R.A., Galle R.F.,
RA George R.A., Lewis S.E., Richards J.R., Ashburner M., Henderson S.N.,
RA Sutton G.G., Wortman J.R., Yandell M.D., Zhang Q., Chen L.X.,
RA Brandon R.C., Rogers Y.-H.C., Blazew R.G., Champe M., Pfeiffer B.D.,
RA Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Miklos G.L.G.,
RA Abril J.F., Agbayani A., An H.-J., Andrews-Pfannkoch C., Baldwin D.,
RA Ballew R.M., Basu P.V., Baxendale J., Bayraktaroglu L., Beasley E.M.,
RA Beeson K.Y., Benos P.V., Berman B.P., Bhandari D., Bolshakov S.,
RA Borkova D., Botchan M.R., Bouck J., Brokstein P., Brattier P.,
RA Burtis K.C., Busan D.A., Butler H., Cadieu E., Center A., Chandra I.,
RA Cherry J.M., Cavley S., Dahlke C., Davenport L.B., Davies P.,
RA de Pablos B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,
RA Dodson K., Dou P., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,
RA Durbin K.J., Evangelista C.C., Ferraz C., Ferrieria S., Fleischmann W.,
RA Foster C., Gabrielian A.E., Garg N.S., Gelbart W.M., Glasser K.,
RA Glodek A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,
RA Harris N.L., Harvey D., Helman T.J., Hernandez J.R., Houck J.,
RA Hostin D., Houston K.A., Howland T.J., Wei M.-H., Ibegwam C.,
RA Jalali M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,
RA Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,
RA Lasko P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,
RA Liu X., Mattei B., McIntosh T.C., McLeod M.P., McPherson D.,
RA Merkulov G., Milshina N.V., Mobarry C., Morris J., Moshrefi A.,
RA Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,
RA Nelson D.R., Nelson K.A., Nixon K., Nusskern D.R., Pacle J.M.,
RA Palazzolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,
RA Reinert K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,
RA Shue B.C., Siden-Kiamos I., Simpson M., Skupski M.P., Smith T.,
RA Spier E., Spradling A.C., Stapleton M., Strong R., Sun E.,
RA Svirskaas R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,
RA Wang Z.-Y., Wassarman D.A., Weinstein G.M., Weissbach J.,
RA Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,
RA Ye J., Yeh R.-F., Zaveri J.S., Zhang M., Zhang G., Zhao Q., Zheng L.,
RA Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,
RA Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;
RT "The genome sequence of Drosophila melanogaster."
RL Science 287:2185-2195(2000).
DR EMBL; AF003436; AAF46120.2;
DR FlyBase; FBgn0029839; CG4660.
DR InterPro; IPR000365; 4HbcoA_thiostrse.
DR Pfam; PF03061; 4HBT; 1.
SQ SEQUENCE 261 AA; 29232 MW; 5AB58960057E3363 CRC64;

Query Match 1.6%; Score 8; DB 5; Length 261;
Best Local Similarity 100.0%; Pred. No. 21;

Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
Qy 211 PELAKWIE 218
|||||
Db 241 PELAKWIE 248

RESULT 15
O24792 PRELIMINARY; PRT; 264 AA.
AC O24792;
DT 01-JAN-1998 (TRENBLrel. 05, Created)
DT 01-JAN-1998 (TRENBLrel. 05, Last sequence update)
DT 01-JUN-2001 (TRENBLrel. 17, Last annotation update)
DE Outer surface protein B (Fragment).
GN OSPB.
OS Borrelia garinii.
OC Bacteria; Spirochaetales; Spirochaetaceae; Borrelia.
OX NCBI_TaxID=29519;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=N333;
RA Fukunaga M., Yabuki M., Nakao M.;
RL Submitted (SEP-1997) to the EMBL/GenBank/DBJ databases.
DR EMBL; AB007099; BAA22341.1; -.
DR HSSP; P14013; IOSP.
DR InterPro; IPR001809; Outrsurface.
DR Pfam; PF00820; Lipoprotein_1; 1.
DR PRINTS; PR00968; OUTRSURFACE.
DR ProDom; PD001127; Outrsurface; 1.
FT NON_TER 264 264
SQ SEQUENCE 264 AA; 28671 MW; 7405E3CEB928EF08 CRC64;

Query Match 1.6%; Score 8; DB 2; Length 264;
Best Local Similarity 100.0%; Pred. No. 21;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 252 DLPLVAED 259
|||||
Db 45 DLPLVAED 52

Search completed: May 8, 2003, 16:44:41
Job time : 99 secs

GenCore version 5.1.4_p5-4578
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: May 8, 2003, 16:25:06 ; Search time 80 Seconds
(without alignments)
807.832 Million cell updates/sec

Title: US-09-926-163B-2

Perfect score: 2516

Sequence: 1 MITRETLKSLPANVOAPPD.....KVILRKTIIEKGVKAAIPA 485

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 908470 seqs, 13250620 residues

Total number of hits satisfying chosen parameters: 908470

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

- A_Geneseq_101002.*
- 1: /SIDS2/gcgdata/geneseq/geneseqp-emb1/AA1980.DAT.*
 - 2: /SIDS2/gcgdata/geneseq/geneseqp-emb1/AA1981.DAT.*
 - 3: /SIDS2/gcgdata/geneseq/geneseqp-emb1/AA1982.DAT.*
 - 4: /SIDS2/gcgdata/geneseq/geneseqp-emb1/AA1983.DAT.*
 - 5: /SIDS2/gcgdata/geneseq/geneseqp-emb1/AA1984.DAT.*
 - 6: /SIDS2/gcgdata/geneseq/geneseqp-emb1/AA1985.DAT.*
 - 7: /SIDS2/gcgdata/geneseq/geneseqp-emb1/AA1986.DAT.*
 - 8: /SIDS2/gcgdata/geneseq/geneseqp-emb1/AA1987.DAT.*
 - 9: /SIDS2/gcgdata/geneseq/geneseqp-emb1/AA1988.DAT.*
 - 10: /SIDS2/gcgdata/geneseq/geneseqp-emb1/AA1989.DAT.*
 - 11: /SIDS2/gcgdata/geneseq/geneseqp-emb1/AA1990.DAT.*
 - 12: /SIDS2/gcgdata/geneseq/geneseqp-emb1/AA1991.DAT.*
 - 13: /SIDS2/gcgdata/geneseq/geneseqp-emb1/AA1992.DAT.*
 - 14: /SIDS2/gcgdata/geneseq/geneseqp-emb1/AA1993.DAT.*
 - 15: /SIDS2/gcgdata/geneseq/geneseqp-emb1/AA1994.DAT.*
 - 16: /SIDS2/gcgdata/geneseq/geneseqp-emb1/AA1995.DAT.*
 - 17: /SIDS2/gcgdata/geneseq/geneseqp-emb1/AA1996.DAT.*
 - 18: /SIDS2/gcgdata/geneseq/geneseqp-emb1/AA1997.DAT.*
 - 19: /SIDS2/gcgdata/geneseq/geneseqp-emb1/AA1998.DAT.*
 - 20: /SIDS2/gcgdata/geneseq/geneseqp-emb1/AA1999.DAT.*
 - 21: /SIDS2/gcgdata/geneseq/geneseqp-emb1/AA2000.DAT.*
 - 22: /SIDS2/gcgdata/geneseq/geneseqp-emb1/AA2001.DAT.*
 - 23: /SIDS2/gcgdata/geneseq/geneseqp-emb1/AA2002.DAT.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	2516	100.0	485	21	AAB23172
2	2143	85.2	485	20	AAU38095
3	802	31.9	503	22	ABG89872
4	802	31.9	503	22	ABG79410
5	419.5	16.7	534	22	AAU39813
6	365.5	14.5	544	23	ABG54978
7	312.5	12.4	788	22	ABG13869
8	221.5	8.8	384	22	AAU35002
9	205	8.1	382	22	AAU34792
10	203	8.1	382	13	AAR28826

11	202.5	8.0	745	22	ABG28573
12	200	7.9	378	22	AAU38095
13	191.5	7.6	384	23	ABG53324
14	190	7.6	366	23	ABG05610
15	181	7.2	368	22	AAU34140
16	181	7.2	368	22	AAU36800
17	162	6.4	391	19	AAW37992
18	162	6.4	391	20	AAW39873
19	162	6.4	391	23	ABG05608
20	133.5	5.3	103	23	ABP33539
21	125.5	5.0	134	18	AAW28071
22	125	5.0	1504	22	ABG93729
23	122.5	4.9	1550	22	ABG20180
24	119	4.7	3685	10	AAU90373
25	114.5	4.6	894	22	ABG59340
26	114	4.5	183	22	ABG25111
27	110.5	4.4	7201	22	ABG71136
28	109.5	4.4	2326	22	ABG50652
29	106	4.2	3685	10	AAU90290
30	105.5	4.2	1560	21	AAW67203
31	105.5	4.2	1561	21	AAW77202
32	105.5	4.2	1562	21	AAW18639
33	105.5	4.2	1562	21	AAW77194
34	105.5	4.2	12199	21	AAW77180
35	105	4.2	1704	16	AAW70188
36	105	4.2	1704	18	AAW34843
37	105	4.2	1704	21	AAW67396
38	105	4.2	1704	22	AAU08938
39	104	4.1	588	17	AAW8209
40	104	4.1	819	22	ABG20721
41	104	4.1	1867	23	ABG66751
42	104	4.1	2059	23	ABP25711
43	103.5	4.1	1533	22	ABG20179
44	103	4.1	395	22	ABG96802
45	103	4.1	877	23	ABG55496

ALIGNMENTS

RESULT 1
AAB23172
ID AAB23172 standard; Protein; 485 AA.
XX
AC AAB23172;
XX
DT 29-JAN-2001 (first entry)
XX
DE Gluconobacter oxydans D-sorbitol dehydrogenase (SLDH).
XX
KW D-sorbitol dehydrogenase; SLDH; L-sorbose production; NADP+ dependent;
KW sorbitol; mannitol; arabitol; oxidation; 2-keto-L-gluconic acid;
KW L-ascorbic acid biosynthesis; vitamin C.
XX
OS Gluconobacter oxydans.
XX
PN WO200055329-A1.
XX
PD 21-SEP-2000.
XX
PF 16-MAR-2000; 2000WO-JP01608.
XX
PR 17-MAR-1999; 99JP-0072810.
XX
PR 06-AUG-1999; 99JP-0224679.
XX
PA (FUJI) FUJISAWA PHARM CO LTD.
XX
PI Shibata T, Ichikawa C, Matsuura M, Noguchi Y, Saito Y;
PI Yamashita M, Takata Y;
XX
DR WPI: 2000-587530/55.
DR N-PSDB: AAA97430.
XX

PT Gluconobacter oxydans derived D-sorbitol dehydrogenase obtainable by
 PT culturing its gene-transformed host cells, useful for producing
 PT L-sorbose or 2-keto-L-gluconic acid as an intermediate in L-ascorbic
 PT acid production -
 XX
 PS Claim 5; Page 58-60; 72pp; Japanese.
 XX
 CC The invention relates to Gluconobacter oxydans D-sorbitol dehydrogenase
 CC (SLDH; AAB23172) and to the gene encoding it (AA97430). SLDH has a
 CC molecular weight of about 54 kDa and catalyses the conversion of
 CC D-sorbitol into L-sorbose using NADP+ as a co-enzyme. SLDH specifically
 CC catalyses the oxidation of sorbitol, mannitol and arabinol, but does not
 CC act on xylitol, ribitol, inositol and glycerol. The invention also
 CC encompasses expression vectors and host cells comprising the
 CC Gluconobacter oxydans SLDH gene, and the recombinant production of SLDH.
 CC The invention further relates to a method for preparing L-sorbose by
 CC contacting the recombinant SLDH with D-sorbitol; a process for producing
 CC 2-keto-L-gluconic acid by contacting recombinantly produced sorbose
 CC dehydrogenase and/or sorbose dehydrogenase with L-sorbose; and a
 CC process for preparing L-ascorbic acid or its alkaline earth metals salts
 CC by conversion from 2-keto-L-gluconic acid. SLDH is useful for producing
 CC L-sorbose or 2-keto-L-gluconic acid as an intermediate in L-ascorbic
 CC acid production. The present sequence represents the Gluconobacter
 CC oxydans SLDH protein.
 XX
 SQ Sequence 485 AA;

Query Match 100.0%; Score 2516; DB 21; Length 485;
 Best Local Similarity 100.0%; Pred. No. 3.9e-221;
 Matches 485; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 1 MITRETLKSLPANVQAPPYDIDGKPGIVHFGVGNFRAHEAFYVQIIEHAPDWAIVG 60
 DB 1 MITRETLKSLPANVQAPPYDIDGKPGIVHFGVGNFRAHEAFYVQIIEHAPDWAIVG 60
 QY 61 GLTGSDRSKKKAEFEKQADCLSLTETAPSGKSTVRVMGALRDYLLAPADPEAVLKLHVD 120
 DB 61 GLTGSDRSKKKAEFEKQADCLSLTETAPSGKSTVRVMGALRDYLLAPADPEAVLKLHVD 120
 QY 121 PAIRIVSMITTEGGYNNINETTGAFLDENAANKADLNKPEKPSVFGYVVEALRRRDAGG 180
 DB 121 PAIRIVSMITTEGGYNNINETTGAFLDENAANKADLNKPEKPSVFGYVVEALRRRDAGG 180
 QY 181 KAFTVMSCDNLRHNGNVARAKFLGAKARDPELAKWIEENATFPNGMVDRIPTVSAEIA 240
 DB 181 KAFTVMSCDNLRHNGNVARAKFLGAKARDPELAKWIEENATFPNGMVDRIPTVSAEIA 240
 QY 241 KKLNAASGLDDDLPLVAEDFHQWLEDFADGRPPLEKAGVQVGVDTWDEYVYKIRMLNA 300
 DB 241 KKLNAASGLDDDLPLVAEDFHQWLEDFADGRPPLEKAGVQVGVDTWDEYVYKIRMLNA 300
 QY 301 GHVMLCFPGILVGYENVDDAIEDSELLGNLKNYLNKNDVIPTLKAPSGMTLEGYRDSVISR 360
 DB 301 GHVMLCFPGILVGYENVDDAIEDSELLGNLKNYLNKNDVIPTLKAPSGMTLEGYRDSVISR 360
 QY 361 FSNKMSDQTLRIASDCGSKVQVFWTETVRRRAIEDKRDLSRIAFAFGIASYLEMLRGRDEKG 420
 DB 361 FSNKMSDQTLRIASDCGSKVQVFWTETVRRRAIEDKRDLSRIAFAFGIASYLEMLRGRDEKG 420
 QY 421 GYVESSEPTYGDAEWKLAKADDFESSLKLPAFDGWRDLDTSELDOQVIVLRLKRIREKGVK 480
 DB 421 GYVESSEPTYGDAEWKLAKADDFESSLKLPAFDGWRDLDTSELDOQVIVLRLKRIREKGVK 480
 QY 481 AAIPIA 485
 DB 481 AAIPIA 485

RESULT 2
 ID AAY49913
 XX AAY49913 standard; Protein; 485 AA.
 AC AAY49913;

XX 27-JAN-2000 (first entry)
 DT Gluconobacter suboxydans L-sorbose reductase protein sequence.
 DE Gluconobacter suboxydans; L-sorbose reductase; genetic engineering;
 XX mutant.
 KW Gluconobacter suboxydans.
 XX AU9920390-A.
 PN 23-SEP-1999.
 PD 11-MAR-1999; 99AU-0020390.
 XX 13-MAR-1998; 98EP-0104546.
 PR (HOFF) HOFFMANN LA ROCHE & CO AG F.
 XX Hoshino T, Tazoe M, Shinjoh M, Kon T;
 PI WPI; 1999-579276/49.
 XX N-PSDB; AAZ35672.
 DR Set of genetically modified mutants not containing L sorbose reductase
 PT
 XX Claim 8; Page 18-21; 33pp; English.
 PS The present invention describes a genetically engineered
 CC microorganism derived from a microorganism belonging to the genus
 CC Gluconobacter or Acetobacter which is characterised in that the
 CC biological activity for reducing L-sorbose is substantially
 CC nullified by gene recombination gene. The present sequence represents
 CC Gluconobacter suboxydans L-sorbose reductase.
 CC N.B. This patent is equivalent to the basic NO9901197 in week 199949.
 XX
 SQ Sequence 485 AA;

Query Match 85.2%; Score 2143; DB 20; Length 485;
 Best Local Similarity 84.5%; Pred. No. 5e-187;
 Matches 410; Conservative 30; Mismatches 45; Indels 0; Gaps 0;
 QY 1 MITRETLKSLPANVQAPPYDIDGKPGIVHFGVGNFRAHEAFYVQIIEHAPDWAIVG 60
 DB 1 MITRETLKSLPANVQAPPYDIDGKPGIVHFGVGNFRAHEAFYVQIIEHAPDWAIVG 60
 QY 61 GLTGSDRSKKKAEFEKQADCLSLTETAPSGKSTVRVMGALRDYLLAPADPEAVLKLHVD 120
 DB 61 GLTGSDRSKKKAEFEKQADCLSLTETAPSGKSTVRVMGALRDYLLAPADPEAVLKLHVD 120
 QY 121 PAIRIVSMITTEGGYNNINETTGAFLDENAANKADLNKPEKPSVFGYVVEALRRRDAGG 180
 DB 121 PAIRIVSMITTEGGYNNINETTGAFLDENAANKADLNKPEKPSVFGYVVEALRRRDAGG 180
 QY 181 KAFTVMSCDNLRHNGNVARAKFLGAKARDPELAKWIEENATFPNGMVDRIPTVSAEIA 240
 DB 181 KAFTVMSCDNLRHNGNVARAKFLGAKARDPELAKWIEENATFPNGMVDRIPTVSAEIA 240
 QY 241 KKLNAASGLDDDLPLVAEDFHQWLEDFADGRPPLEKAGVQVGVDTWDEYVYKIRMLNA 300
 DB 241 KKLNAASGLDDDLPLVAEDFHQWLEDFADGRPPLEKAGVQVGVDTWDEYVYKIRMLNA 300
 QY 301 GHVMLCFPGILVGYENVDDAIEDSELLGNLKNYLNKNDVIPTLKAPSGMTLEGYRDSVISR 360
 DB 301 GHVMLCFPGILVGYENVDDAIEDSELLGNLKNYLNKNDVIPTLKAPSGMTLEGYRDSVISR 360
 QY 361 FSNKMSDQTLRIASDCGSKVQVFWTETVRRRAIEDKRDLSRIAFAFGIASYLEMLRGRDEKG 420
 DB 361 FSNKMSDQTLRIASDCGSKVQVFWTETVRRRAIEDKRDLSRIAFAFGIASYLEMLRGRDEKG 420
 QY 421 GYVESSEPTYGDAEWKLAKADDFESSLKLPAFDGWRDLDTSELDOQVIVLRLKRIREKGVK 480
 DB 421 GYVESSEPTYGDAEWKLAKADDFESSLKLPAFDGWRDLDTSELDOQVIVLRLKRIREKGVK 480

PR 27-AUG-1999; 99DE-1040765.
PR 31-AUG-1999; 99US-0151572.
PR 03-SEP-1999; 99DE-1042076.
PR 03-SEP-1999; 99DE-1042079.
PR 03-SEP-1999; 99DE-1042086.
PR 03-SEP-1999; 99DE-1042087.
PR 03-SEP-1999; 99DE-1042088.
PR 03-SEP-1999; 99DE-1042095.
PR 03-SEP-1999; 99DE-1042123.
PR 03-SEP-1999; 99DE-1042125.
XX (BADI) BASF AG.
XX
XX Pompejus M, Kroeger B, Schroeder H, Zelder O, Haberhauer G;
XX
XX WPI: 2001-061975/07.
XX N-PSDB; AAF71527.
XX
XX New isolated Corynebacterium glutamicum nucleic acid encoding a sugar
XX metabolism and oxidative phosphorylation protein for production or
XX modulation of production of fine chemicals e.g. amino acids,
XX carbohydrates or enzymes
XX
XX Claim 20; Page 615-616; 1246pp; English.
XX
XX AAF71360 to AAF71750 encode the Corynebacterium glutamicum sugar
XX metabolism and oxidative phosphorylation (SMP) proteins given in
XX AAB79243 to AAB 79633 which are involved in carbon metabolism and
XX energy production. The C. glutamicum SMP gene can be used in vectors
XX (II) for expression in host cells and production or modulation of
XX production of fine chemicals, such as, an organic acid, a proteinogenic
XX or nonproteinogenic amino acid (preferred), a purine or pyrimidine base,
XX a nucleoside, a nucleotide, a lipid, a saturated or unsaturated fatty
XX acid, a diol, a carbohydrate, an aromatic compound, a vitamin, a
XX cofactor, a polyketide, or an enzyme. The presence of (I) or SMP proteins
XX (III) encoded by them are used for diagnosing the presence or activity of
XX Corynebacterium diptheriae in a subject. (I), (II), (III) or host cells
XX containing them are used to map genomes of organisms related to
XX C. glutamicum, identify and localise C. glutamicum sequences of interest,
XX in evolutionary studies, in determining SMP protein regions required
XX for function, in modulating SMP protein activity, in modulating the
XX metabolism of sugars, and in modulating high-energy molecule production
XX in a cell (i.e. ATP, NADPH).
XX
XX Sequence 503 AA:
Query Match 31.9%; Score 802; DB 22; Length 503;
Best Local Similarity 39.0%; Pred. No. 2.3e-64;
Matches 168; Conservative 82; Mismatches 163; Indels 18; Gaps 6;
QY 2 ITRETLKSL--PANYQAPPYDIDGKPGIVHFGVGNFFRAHEAFYVEQILEH--APDWAI 57
Db 7 LNTENLOETASTSGVQIPAFNADVAPGIVHFGVGFHRAHQAMYLNEMNGKALDWCI 66
QY 58 VGVGLTGSRSKKAEEFAQDCLSLTETAPSGKSTVRVMGALRDYLLAPADPEAVLKH 117
Db 67 IGMGVMPDSVRMDA--LASQDHLVTLTKAPDGLTDQKIISIDYFAPEDPARAVAT 124
QY 118 LVDPAIRIVSMITTEGGYNNITTCGAFLENAAVKADLK-----NPEKSTVFGYVEAL 172
Db 125 LAQDSIRVSLVTVTEGGYNIDPATEDFDHNPRIVADREALQAGDTSTLTQTFGLITAAAL 184
QY 173 RRRWDAGGKAFVMSGDLNRHNGNARKAFGLYAKARDELPALAKWTEENATFPNGMVDRT 232
Db 185 ISRKESGSTPFTIMSCDNIQGNGLAKREFLAFAHSVSSELEWENNVAFFNSMVDRT 244
QY 233 PTVS---AFIAKLNAAAGLDDLPVLEDFHQLWLEQFADGRPLEKAGVQMVGVDVT 289
Db 245 PETTDGDRDDKEI---GYIDAWPVVSEDTQWVLEDAFTQCPAYEEVGQVVSDEVP 300
QY 290 WEVVKIRMLNAGHVMLCFPGILVGVENVDDATEDSELNGLNKNLYNKDVIPTLKAPSGMT 349
Db 301 YELMKRLNLNASHOGLCYFGHLAGHHMVDHVMADTRFQDFLLAYMEREATPTLKLPGVD 360

QY 350 LEGYRDSVLSRESNKAMSDQTLRIASDGCCKYQVFWETVRRRAIEDKRDLSRTAFGIASY 409
Db 361 LDAYRRQLIAREGNAAKVDTPRLCAESSDRIPKWLPPVVRNLAAGRDVTLSSAIVASW 420
QY 410 LEMLRGRDEKG 420
Db 421 ARYAEGTDEQG 431
RESULT 5
AAU39813
ID AAU39813 standard; Protein; 534 AA.
XX
XX AC AAU39813;
XX DT 13-FEB-2002 (first entry)
XX
XX Propionibacterium acnes immunogenic protein #709.
DE
XX SAPHO syndrome; synovitis; acne; pustulosis; hypertosis; osteomyelitis;
KW uveitis; endophthalmitis; bone; joint; central nervous system; ELISA;
KW inflammatory lesion; acne vulgaris; enzyme linked immunosorbent assay;
KW dermatological; osteopathic; neuroprotectant.
XX
XX Propionibacterium acnes.
XX WO200181581-A2.
XX
XX PD 01-NOV-2001.
XX
XX PF 20-APR-2001; 2001WO-US12865.
XX
XX PR 21-APR-2000; 2000US-199047P.
PR 02-JUN-2000; 2000US-208841P.
PR 07-JUL-2000; 2000US-216747P.
XX
XX (CORI-) CORIXA CORP.
XX
XX Skeiky YAM, Persing DH, Mitcham JL, Wang SS, Bhatia A;
PI L'maisonneuve J, Zhang Y, Jen S, Carter D;
XX
XX WPI: 2001-616774/71.
DR N-PSDB; AAS59508.
XX
XX Propionibacterium acnes polypeptides and nucleic acids useful for
PT vaccinating against and diagnosing infections, especially useful for
PT treating acne vulgaris -
XX
XX Claim 4; SEQ ID No 1008; 1069pp; English.
XX
XX Sequences AAU39105-AAU68017 represent Propionibacterium acnes immunogenic
CC polypeptides. The proteins and their associated DNA sequences are used in
CC the treatment, prevention and diagnosis of medical conditions caused by
CC P. acnes. The disorders include SAPHO syndrome (synovitis, acne,
CC pustulosis, hypertosis and osteomyelitis), uveitis and endophthalmitis.
CC P. acnes is also involved in infections of bone, joints and the central
CC nervous system, however it is particularly involved in the inflammatory
CC lesions associated with acne vulgaris. A method for detecting the
CC presence or absence of P. acnes in a patient comprises contacting a
CC sample with a binding agent that binds to the proteins of the invention
CC and determining the amount of bound protein in the sample. The
CC polypeptides may be used as antigens in the production of antibodies
CC specific for P. acnes proteins. These antibodies can be used to
CC downregulate expression and activity of P. acnes polypeptides and
CC therefore treat P. acnes infections. The antibodies may also be used as
CC diagnostic agents for determining P. acnes presence, for example, by
CC enzyme linked immunosorbent assay (ELISA).
CC Note: The sequence data for this patent did not form part of the printed
CC specification, but was obtained in electronic format directly from WIPO
CC at ftp.wipo.int/pub/published_pct_sequences.
XX
XX Sequence 534 AA;

CC nucleic acid sequence is useful in the detection and/or amplification of
CC nucleic acid sequence, particularly to identify *Lactococcus lactis* or
CC related species. The proteins of the invention are useful for the
CC biosynthesis or biodegradation of a composition of interest. The
CC invention helps research in lactic bacteria, particularly useful in the
CC production of yogurt and cheese.
CC Note: The sequence data for this patent is based on equivalent patent
CC WO200177334 (published 18-Oct-2001) which is available in electronic
CC format directly from WIPO at ftp.wipo.int/pub/published_pct_sequences.

Query Match	14.5%	Score 365.5;	DB 23;	Length 544;
Best Local Similarity	26.4%;	Pred. No. 2.le-24;		
Matches 132;	Conservative 84;	Mismatches 213;	Indels 71;	Gaps 16;

QY	1	MITRETLSLPANQAPPYDIDGI-----KPGIVHETGVGNFFRAHAFAFYVEQIILEHAPD	54
DB	9	LIKGEFPK--AKIKIPNFODKVIESTADNPVWVHGGGJLFCFHVAVAQDILLNQOGL	66
QY	55	WAIYGVGLTSDRSKKAAEFPKADCLYSLTETAPSGKSTVRVMGALRDYLLAPADPEAV	114
DB	67	NSGLIVAETVDEVIDKI--YRAYNNRFLSYTMKSDGTFDKELIASVAESFYFNAENTEG	124
QY	115	LKHLV----DPAIRVSMTEITGGYININETGAEFDLENAAYKADLNKPEKSTVFYGYVE	170
DB	125	WAHLTKVFENPSLOLTFSTIEKYSLRDSK---DQLTPLALEDQMGRSKPKTNMGAVTY	181
QY	171	ALRRWDAGGKAFVMSCDNLRHNGNVARKAFGLYAKARDPELAKWIE-----	218
DB	182	LLYARFAGKRPPIAMVSTDNFSENGLKQEAILTAKA-----WVNEIVEODFYDYL	234
QY	219	---ENATFPNGMVDRIPTVSAEIAKLINAAAGLDDDLPLVA-----EDFHOW	263
DB	235	INPKKVSFPNWSMIDRITPNPSEKVAKULL-TADGFEDTEILHSQKHTNAPFGNTEEVHYL	293
QY	264	VLEDOQDAGRPLPEKAGVQWVG--DVTDWYVKI-RMLNACHVMVLCPPGILVGVENVDDA	320
DB	294	VIEDAFNNGRALEKSGVILTDRTEVNDADQMKVTACLNPLHTLALFGLSLDHYHSIWE	353
QY	321	IEDSELLGNLKNLYNKDVIPTLKAPSGMTLBYRDSVLS-RFSNKAUSDQTLRTASDGS	379
DB	354	VANPDLLALIKNLGYGALPVVKNPKIINPKDFIDQLLTKRLPNKKNIPDTPORTAADTSQ	413
QY	380	KVQVFWTETVRRATEDKR---DLSRIAFTAGSYALEMLRG-RDEKGGTYESESSEPTYGDAE	434
DB	414	KIPRYGVTIHGYIANPRFVSKELEFPPLVTAANCRYLIGINDELESFSPSPDPL-----	468
QY	435	WKLAKADDFESSLKLPAFDG	454
DB	469	--LEELQAFVADVKL--DG	483

RESULT 7	
ABG13869	
ID	ABG13869 standard; Protein; 788 AA.
XX	
XX	ABG13869;
XX	
XX	18-FEB-2002 (first entry)
DT	
XX	
XX	Novel human diagnostic protein #13860.
DE	
XX	
KW	Human; chromosome mapping; gene mapping; gene therapy; forensic;
KW	food supplement; medical imaging; diagnostic; genetic disorder.
XX	
XX	OS Homo sapiens.
XX	
XX	WO200175067-A2.
PN	
XX	
XX	11-OCT-2001.
PD	
XX	
XX	30-MAR-2001; 2001WO-US08631.
PF	

```
XX 31-MAR-2000; 2000US-0540217.
PR 23-AUG-2000; 2000US-0649167.
XX (HYSE-) HYSEQ INC.
PA Drmanac RT, Liu C, Tang YT;
PI WPI: 2001-639362/73.
DR N-PSDB; AAS78056.
XX New isolated polynucleotide and encoded polypeptides, useful in
PT diagnostics, forensics, gene mapping, identification of mutations
PT responsible for genetic disorders or other traits and to assess
PT biodiversity
XX Claim 20; SEQ ID No 44228; 103pp; English.
XX The invention relates to isolated polynucleotide (I) and
XX polypeptide (II) sequences. (I) is useful as hybridisation probes,
XX polymerase chain reaction (PCR) primers, oligomers, and for chromosome
XX and gene mapping, and in recombinant production of (II). The
XX polynucleotides are also used in diagnostics as expressed sequence tags
XX for identifying expressed genes. (I) is useful in gene therapy techniques
XX to restore normal activity of (II) or to treat disease states involving
XX (II). (II) is useful for generating antibodies against it, detecting or
XX quantitating a polypeptide in tissue, as molecular weight markers and as
XX a food supplement. (II) and its binding partners are useful in medical
XX imaging of sites expressing (II). (I) and (II) are useful for treating
XX disorders involving aberrant protein expression or biological activity.
XX The polypeptide and polynucleotide sequences have applications in
XX diagnostics, forensics, gene mapping, identification of mutations
XX responsible for genetic disorders or other traits to assess biodiversity
XX and to produce other types of data and products dependent on DNA and
XX amino acid sequences. AB00010-AB030377 represent novel human
XX diagnostic amino acid sequences of the invention.
XX Note: The sequence data for this patent did not appear in the printed
XX specification, but was obtained in electronic format directly from WIPO
XX at ftp.wipo.int/pub/published_pct_sequences.
XX Sequence 788 AA;
SQ
Query Match 12.4%; Score 312.5; DB 22; Length 788;
Best Local Similarity 37.6%; Pred. No. 2.6e-19;
Matches 67; Conservative 37; Mismatches 69; Indels 5; Gaps 3;
QY 255 LVAEDFHQVLEDOFADGRPPLEKAGQVGVDTWDEYVKIRMLNAGHVMLCFPGTLVGY 314
DB 3 VACEPRQWVIEDNFVAGRPWEKAGAEVLSDVLPYEMKRLMLNGSHSLAYLGLYLAGY 62
QY 315 ENVVDIAIEDSELGNLKNYLNDKVIPTLKAPSGMTLEGYRDSVISRFSNKMSDQTLRIA 374
DB 63 OHINDCMEDEHYRYAAYGLMLQEQAPTLAV-QGVDLQDYANRLIARYSNPALRHRWTQIA 121
QY 375 SDGCSKVQVFWETVR--RAIEDKRDLRIAFGIASYLEMLGRDEKGYTESSEPTY 430
DB 122 MDGSKQLPQRMLDVSRWHLAHDSKFDL--LALGVAGWMRYGVGVGQNPFIETSDPLF 177
RESULT 8
AAU35002
ID AAU35002 standard; Protein; 384 AA.
XX
AC AAU35002;
XX
DT 14-FEB-2002 (first entry)
DE Enterococcus faecalis cellular proliferation protein #289.
XX
KW Antisense; prokaryotic cellular proliferation protein;
KW antibiotic; antibacterial; drug design.
XX
OS Enterococcus faecalis.
XX
XX WO200170955-A2.
XX 27-SEP-2001.
XX
XX 21-MAR-2001; 2001WO-US09180.
XX
XX 21-MAR-2000; 2000US-191078P.
XX 23-MAY-2000; 2000US-206848P.
XX 26-MAY-2000; 2000US-207727P.
XX 23-OCT-2000; 2000US-242578P.
XX 27-NOV-2000; 2000US-253625P.
XX 22-DEC-2000; 2000US-257931P.
XX 16-FEB-2001; 2001US-269308P.
XX (ELIT-) ELITRA PHARM INC.
PA
XX
XX Haselbeck R, Ohlsen KL, Zyskind JW, Wall D, Trawick JD, Carr GJ;
XX Yamamoto RT, Xu HH;
XX
XX WPI: 2001-611495/70.
XX N-PSDB; AAS52861.
XX
XX New polynucleotides for the identification and development of
XX antibiotics, comprise sequences of antisense nucleic acids -
XX
XX Example 3; Seq ID No 10595; 511pp; English.
XX
XX The invention relates to antisense inhibitors of genes essential to
XX prokaryotic cellular proliferation, their use in identifying the
XX genes, their use in the discovery of novel antibiotics, the essential
XX genes themselves and the encoded proteins. The prokaryotes used are
XX Escherichia coli, Staphylococcus aureus, Salmonella typhi, Klebsiella
XX pneumoniae, Pseudomonas aeruginosa and Enterococcus faecalis. The
XX invention is also useful for the identification of potential new targets
XX for antibiotic development. The antisense nucleic acids can also be used
XX to identify proteins used in proliferation, to express these proteins,
XX and to obtain antibodies capable of binding to the expressed proteins.
XX The proteins can be used to screen compounds in rational drug discovery
XX programmes. The antisense nucleic acid sequence is also useful to screen
XX for homologous nucleic acids which are required for cell proliferation in
XX a wide variety of organisms. The present sequence represents an
XX essential prokaryotic cellular proliferation protein.
XX Note: The sequence data for this patent did not form part
XX of the printed specification, but was obtained in electronic
XX format directly from WIPO at
XX ftp.wipo.int/pub/published_pct_sequences.
XX
XX Sequence 384 AA;
SQ
Query Match 8.8%; Score 221.5; DB 22; Length 384;
Best Local Similarity 22.6%; Pred. No. 1.8e-11;
Matches 83; Conservative 61; Mismatches 123; Indels 101; Gaps 14;
QY 29 VIFGVGNFRAHEAFYVEQILEHAPDWAIVGLTGSDRSKKAEFEKADQCLYSLTETA 88
DB 4 VHFAGNIGRG---FIGEIL-----AKNGFHITFVD-----VNET- 35
QY 89 PSGKSTVRVMGALRDYLLAPADPEAVLKHVDPATRIVSMITTEGYNINETTGAFDLEN 148
DB 36 -----IIQALKERKSYTIELADAS-----HQQINVENT-----GLNNTEPEKVVE 77
QY 149 AAVKADLKNPEKPSVTFVGVVVEALRRRRDAGKA-----FTVMSCDNLRHNGNVARKAF 203
DB 78 ATAEDLVTTAIGPNILPRIAELIAQIGIDARAEANCQKPLDIIACENM-----IGGSTFL 132
QY 204 GYAKARDEPELAKWIEENA-----TFPQWVDRIPTVSAETAKKLNAAAGDLDPLVA 257
DB 133 A-----EEVAKYLNPAVAEQWIGFPDAADVIRVPLQKHE-----DPLFVGQV 174
QY 258 EDFHOWVLEDOFADGRPPLEKAGQVGVDTWDEYVKIRMLNAGHVMLCFPGTLVGYENV 317
DB 175 EPFCEWIDDTNRKAK-EIQLEGVHYVADLEPYIERKLFVSNTGATVAYTGALLGYQTI 233
```



```
PS Claim 22; Page 47 + 36-37; 55pp; English.
XX
XX Two separate genes are capable of inducing novel polyol biosynthesis
CC in transgenic plants. One gene is bacterial in origin, e.g. E. coli
CC derived mtd encoding mannitol-1-p dehydrogenase (AAQ31199). The gene
CC for this enzyme has been described by Lee and Sailer, J. Bact., 153:2,
CC 685-692 (1983). The other gene is lmi encoding myo-inositol O-
CC methyl transferase, derived from a stress tolerant plant (AAQ31200).
XX
XX Sequence 382 AA;
SQ
Query Match 8.1%; Score 203; DB 13; Length 382;
Best Local Similarity 21.3%; Pred. No. 8.9e-10;
Matches 101; Conservative 75; Mismatches 160; Indels 138; Gaps 20;
QY 29 VHFVGNGFFRAHEAFYVEQILEHAFDWAIVGVGLGSDRSBKKKAEFFKAQDC--LYSLTE 86
DB 4 LHFAGAGTGR-----FIGKLADA-----GIQLTFADVQNVVLDAIARHYSQVHVGE 53
QY 87 T----APSGKSTVRVMGALRDYLLAPADPEAYVLKHLVDPAIRIVSMITTEGGYNINETT 142
DB 54 TEQVDVFGVNAVSSIGDDVDVLIQVD--LVTNRVGPV----- 91
QY 143 AFDLENAVAKDLNPKRSTVFGIVVEALRRRWAGGKAFVTVMSCDNLRHNGNVARKAF 202
DB 92 ----LERIA-----PAIAKGQV---KRKEQGNESPLNIIACENN--VRGTTQLK-- 131
QY 203 LGYAKARDPELAK--WTEENATPENGVDRTTPTVSAETIAKLNAAAGLDDDLPLVAEDPH 261
DB 132 -GHVNALPEDAKAWEEHVGFDNAVRIVP-----SASATNDPLEVTVETFS 180
QY 262 QWVLE--DQFADGRPPLEKAGVQMGVDVTWVEYVKIRMLNAGHVMLCFPGILVGVENVDDA 320
DB 181 EWIVDKTEFGALPNI--PGMELTDNLMAFVERKLTTLTNGHATAYLGLKLAGHQTIRDA 238
QY 321 IEDSELLGNLKNVLDVPTLKAQSGWTL-----EGYRDSVLSRFSNKAMSDQ 369
DB 239 ILDEKIRAVVKGAMEB-----SGAVLIKRYGFDADKHAAYIQKILGRFENPYLKDD 289
QY 370 TLRASDGCQKVQVFWTETVRRRAIEDKRLSRIAFGIAVLEMLRGRDEKGYTESSEPT 429
DB 290 VERVG-----RQPLRLKSAGDRLIKPL-----GTLEWGLPH 321
QY 430 YGDAEWKLAKADDFESSLKLPADFGWRDLDTSELQKVIVLRKIITREKGVKAAI 483
DB 322 KNRIE--GIAAAMHFR-----SEDDPQAQELAALIAADKGPQAAL 358
RESULT 11
ABG28573
ID ABG28573 standard; Protein; 745 AA.
XX
XX ABG28573;
XX
XX 18-FEB-2002 (first entry)
XX
XX Novel human diagnostic protein #28564.
XX
XX Human; chromosome mapping; gene mapping; gene therapy; forensic;
XX food supplement; medical imaging; diagnostic; genetic disorder.
XX
XX Homo sapiens.
XX
XX WO200175067-A2.
XX
XX 11-OCT-2001.
XX
XX 30-MAR-2001; 2001WO-US08631.
XX
XX 31-MAR-2000; 2000US-0540217.
XX
XX 23-AUG-2000; 2000US-0649167.
XX
XX (HYSE-) HYSEQ INC.
PA
Drmanac RT, Liu C, Tang YT;
WPI: 2001-639362/73.
N-PSDB: AAS92760.
New isolated polynucleotide and encoded polypeptides, useful in
diagnostics, forensics, gene mapping, identification of mutations
responsible for genetic disorders or other traits and to assess
biodiversity
Claim 20; SEQ ID NO 58932; 103pp; English.
The invention relates to isolated polynucleotide (I) and
polypeptide (II) sequences. (I) is useful as hybridisation probes,
polymerase chain reaction (PCR) primers, oligomers, and for chromosome
and gene mapping, and in recombinant production of (II). The
polynucleotides are also used in diagnostics as expressed sequence tags
for identifying expressed genes. (I) is useful in gene therapy techniques
to restore normal activity of (II) or to treat disease states involving
(II). (II) is useful for generating antibodies against it, detecting or
quantitating a polypeptide in tissue, as molecular weight markers and as
a food supplement. (II) and its binding partners are useful in medical
imaging of sites expressing (II). (I) and (II) are useful for treating
disorders involving aberrant protein expression or biological activity.
The polypeptide and polynucleotide sequences have applications in
diagnostics, forensics, gene mapping, identification of mutations
responsible for genetic disorders or other traits to assess biodiversity
and to produce other types of data and products dependent on DNA and
amino acid sequences. ABG00010-ABG30377 represent novel human
diagnostic amino acid sequences of the invention.
Note: The sequence data for this patent did not appear in the printed
specification, but was obtained in electronic format directly from WIPO
at ftp.wipo.int/pub/published_pct_sequences.
SQ Sequence 745 AA;
Query Match 8.0%; Score 202.5; DB 22; Length 745;
Best Local Similarity 30.8%; Pred. No. 2.8e-09;
Matches 45; Conservative 30; Mismatches 70; Indels 1; Gaps 1;
QY 283 MYGDTDYEVYKIRMLNAGHVMLCFPGILVGVENVDDATEDSELLGNLKNVLDVPTL 342
DB 1 MYNDLPWPEEMKRLMLNGSHSLAYLGLYLSGFAHISDCQDRAFRHARTLMDLDEQAPT 60
QY 343 KAPSGMTLEGYRDSVISRFSNKAMSDQTLRIASDGCQKVQVFWTETVRRRAIEDKRLSRI 402
DB 61 QI-KVDVLTQYADKLIARFANPALKHKTQWIAMDGSKLPQRLMAGIRITHOGRDWSLL 119
QY 403 AFGIASYLEMLRGRDEKGGTYESSEP 428
DB 120 ALGVAGWMRYVSGVDDAGNAIDVRDP 145
RESULT 12
AAU38095
ID AAU38095 standard; Protein; 378 AA.
XX
XX AAU38095;
XX
XX 14-FEB-2002 (first entry)
XX
XX Streptococcus pneumoniae cellular proliferation protein #524.
XX
XX Antisense; prokaryotic cellular proliferation protein;
XX antibiotic; antibacterial; drug design.
XX
XX Streptococcus pneumoniae.
OS
XX WO200170955-A2.
XX
XX 27-SEP-2001.
XX
```


Db 137 VSDYLSSEKLYLSKFI-----GPPNAADVIRP-----AQKHDKVLYVEVEPF 180
Qy 261 HQMWLEQFADGRPPLEKAGVQMGVDVTDWEYVKIRMLNAGHVMLCFPGILVGYENVDDA 320.
Db 181 SEMVIDASHLKKN-EIKLEGVHYTTDLEPFIERKUFVSNGHAAVAYSAYGKYKILLEG 239
Qy 321 IEDSELLGNLK-----NYLNKDVITPLKAPSGMTLEGYRDSVISRFSNKA 365
Db 240 LQDEEILNLIKAVOKETRALLLAKWAQYKQD-----ELIKYHELLIISRFSNPE 288
Qy 366 MSDQTLRIASDGCQVQVFWETVTRRAIEDKRD 398
Db 289 IIDEVSRVARTPIRKLGG--YDERFIRPIRELND 319
RESULT 14
ID ABB05610 standard; Protein; 366 AA.
AC ABB05610;
XX
DT 24-APR-2002 (first entry)
XX
DE B. subtilis mannitol-1-phosphate dehydrogenase protein SEQ ID NO:28.
KW Aspergillus oryzae; mutant; modified; reduced transcription; hormone;
KW reduced translation; reduced secretion; receptor; antibody; reporter;
KW enzyme; lipase; mannitol-1-phosphate dehydrogenase.
XX
OS Bacillus subtilis.
OS Synthetic.
XX
XX US6323002-B1.
XX
PD 27-NOV-2001.
XX
XX 25-JUN-1999; 99US-0339972.
XX
PR 12-SEP-1997; 97US-0928692.
PR 13-SEP-1996; 96US-0713312.
XX
XX (NOVO) NOVOZYMES BIOTECH INC.
XX
XX Brody H, Yaver DS, Lamsa M, Hansen K;
XX
DR WPI; 2002-163017/21.
XX
PT Producing a polypeptide using a cell for reducing the production of the
PT polypeptide, comprises inserting DNA into the genome of the cell at a
PT position not within the polypeptide coding sequence or a regulatory
PT sequence
XX
PS Example 15; Column 119-122; 129pp; English.
XX
CC The present invention describes a method for producing a polypeptide (P1)
CC comprising cultivating a mutant cell whose parent cell comprises a DNA
CC sequence encoding P1, by introducing a nucleic acid construct into the
CC genome of the parent cell at a locus not within the P1 sequence, so
CC that P1 transcription, translation or secretion is reduced, and
CC recovering P1. The method is used to produce a polypeptide, such as a
CC recombinant or heterologous hormone, hormone variant, receptor, antibody,
CC reporter or enzyme, particularly an oxidoreductase, transferase,
CC hydrolase, lyase, isomerase or ligase. The present sequence represents
CC a Bacillus subtilis mannitol-1-phosphate dehydrogenase protein which has
CC 34.7% identity with mutant Aspergillus oryzae DEBY932, which is used in
CC an example from the present invention.
XX
SQ Sequence 366 AA;

Query Match 7.6%; Score 190; DB 23; Length 366;
Best Local Similarity 22.5%; Pred. No. 1.3e-08;
Matches 67; Conservative 50; Mismatches 121; Indels 60; Gaps 10;

Qy 29 VHFVGNGFFRAHEAFYVQILEHAPDWAIVGVLGTSRDKKKAEEFKAOCLYSLTETA 88
Db 4 LHFAGNIGRG---FIGALLHHS-----GYDVVF-----A 30
Qy 89 PSGKSTVRVMGALRDYLLAPADPEAVLKLHVDPAIRIVSMITTEGGYINNETTGAFDLEN 148
Db 31 DYNETMVSLLNEKKEYTVELAE-EGRSSEIIGPVSAINSQSQTIELYRL-----MNE 81
Qy 149 AAVKADLANPEKPSIVFGYVVEALRRRRDAGKAFVMSCDNLRHNGNVARKAFLYAKA 208
Db 82 AALITTAAGPNVNLKLIAPSIAGLRRNTA--NTLNIACENMIGGSSFLKKEI--YSHL 137
Qy 209 RDELAKWIEENATPPNGMVDRIPTVSAETAKKLNAAAGLDDDDLPVLAEDFHOWVLEDO 268
Db 138 TEAE-QKSVSETLGFNSAVDRIVPIQHHE-----DPLKVSVEFFEHVIDES 184
Qy 269 FADGRPPLEKAGVQMGVDVTDWEYVKIRMLNAGHVMLCFPGILVGYENVDDAIEDSEL 326
Db 185 GFKGKTPVIN-GALFVDDLTPTVIERKLTFTVNTGHAVTAYVGVQKGLTKVKEAIDHPEI 241
RESULT 15
ID AAU34140 standard; Protein; 368 AA.
AC AAU34140;
XX
DT 14-FEB-2002 (first entry)
XX
DE Staphylococcus aureus cellular proliferation protein #416.
KW Antisense; prokaryotic cellular proliferation protein;
KW antibiotic; antibacterial; drug design.
XX
OS Staphylococcus aureus.
XX
PN WO200170955-A2.
XX
PD 27-SEP-2001.
XX
PF 21-MAR-2001; 2001WO-US09180.
XX
PR 21-MAR-2000; 2000US-191078P.
PR 23-MAY-2000; 2000US-206848P.
PR 26-MAY-2000; 2000US-20727P.
PR 23-OCT-2000; 2000US-242578P.
PR 27-NOV-2000; 2000US-253625P.
PR 22-DEC-2000; 2000US-257931P.
PR 16-FEB-2001; 2001US-269308P.
XX
PA (ELIT-) ELITRA PHARM INC.
XX
XX Haselbeck R, Ohlsen KL, Zyskind JW, Wall D, Trawick JD, Carr GJ;
PI Yamamoto RT, Xu HH;
PI
XX WPI; 2001-611495/70.
DR N-PSDB; AAS51999.
XX
XX New polynucleotides for the identification and development of
PT antibiotics, comprise sequences of antisense nucleic acids -
XX
PS Example 3; Seq ID No 5636; 511pp; English.
XX
CC The invention relates to antisense inhibitors of genes essential to
CC prokaryotic cellular proliferation, their use in identifying the
CC genes, their use in the discovery of novel antibiotics, the essential
CC genes themselves and the encoded proteins. The prokaryotes used are
CC Escherichia coli, Staphylococcus aureus, Salmonella typhi, Klebsiella
CC pneumoniae, Pseudomonas aeruginosa and Enterococcus faecalis. The
CC invention is also useful for the identification of potential new targets
CC for antibiotic development. The antisense nucleic acids can also be used
CC to identify proteins used in proliferation, to express these proteins,

CC and to obtain antibodies capable of binding to the expressed proteins.
CC The proteins can be used to screen compounds in rational drug discovery
CC programmes. The antisense nucleic acid sequence is also useful to screen
CC for homologous nucleic acids which are required for cell proliferation in
CC a wide variety of organisms. The present sequence represents an
CC essential prokaryotic cellular proliferation protein.
CC Note: the sequence data for this patent did not form part
CC of the printed specification, but was obtained in electronic
CC format directly from WIPO at
CC ftp.wipo.int/pub/published_pct_sequences.
XX
XX

SQ Sequence 368 AA;
Query Match 7.2%; Score 181; DB 22; Length 368;
Best Local Similarity 21.4%; Pred. No. 8.6e-08;
Matches 99; Conservative 64; Mismatches 168; Indels 132; Gaps 17;
QY 29 VHFVGNFRAHEAYVEOILEHAPDWAIVGVLGTGSDRSKKKAEFKAQDCLYSLTETA 88
DB 4 VHFAGNIGRG----FIGYIL-----ADNNVKVTFADVNEEIIINAL-AHDHQYDVI-LA 51
QY 89 PSGKSTVRVMGALRDYLLAPADPEAVLKHLDVPAIRIVSMTITTEGGYNINETTGAFDLEN 148
DB 52 DESKTTTRV-----NNVDA-----INSMQPSEALKQ 77
QY 149 AAVKADLNKPEKSTVFGYVVEALRRRRWDAGGKFTVMSCDNLRHNGNVARKAFLGYAKA 208
DB 78 AILEADIITTAGVGNILPIIAKSFAPFLKEKTNHNVIVACENAIMATDLKKAVL----- 132
QY 209 RDELAKEWTEENATPFGMWDRITPTVSAEIAKLNAAAGLDDDLPLVAEDFHWQWLEDO 268
DB 133 ----DITGLGHNIHFANSADVRIPLQKNE-----NILDVMVEPEYEWVVEKD 177
QY 269 FADGRPLEKAGVQWGDVTDWEYVKIRMLNAGHVMLCPGILVGVENVDDAIEDSELLG 328
DB 178 ANYG-PELNH--IKYVDDLTPYIERKLLTVNTGHAYLAGFAKATVLDRAVKDSSIEA 234
QY 329 NLKNYLKNDVIPTLKAPSGMTLE-----GYRDSVISRFSNKAMSDQTLRIASDGCS 379
DB 235 GLRRVL-----AETSQVITNEFDTEAEQAGYVEKIIDRENNSVLSDEVTRVGRG--- 284
QY 380 KQVQFWTEVTRRAIEDKRDLRSIAFGIASYLEMLRGD-EKGTYESSEPTYGDAEWKLA 438
DB 285 -----TLRKIGPKDR-----IKPLKLYLNKDLERTGLLNTAA-----LLL 320
QY 439 KADDFESSLKLPAFDGWRDLDTSELDOKVIVLKRKIIREKGVKA 481
DB 321 KYDD-----TADQETVEKNNYIKEHGLKA 344

Search completed: May 8, 2003, 16:32:26
Job time : 90 secs

THIS PAGE BLANK (USPTO)


```
Db 4 LHFAGNIGRG-----FIGKLADA-----GIQTFADVNVQVLDALNARH----- 44
Qy 89 PSCKSTVRVMGALRDYLLAPADPEAVLKLHVDPAIRIVSMITTEGGYININETTGAFDLEN 148
Db 45 ---SYQVHVVGTEQ-----VD-----TVSGVNAVSGIGD-DVVD 75
Qy 149 AAVKADLNKPEKPSVFGVYVEAL-----RRWDAGGKAFVTMSCDNLRHNGNVARKAPL 203
Db 76 LIAQVDLVTRVGVVLERIAPAIKAGQVKRKEQGNESPLNIIACENM-VRGTTQLK--- 131
Qy 204 GYAKARDEPELAK-WIEENATFNGWVDRTPTVSAEIAKLNAAAGLDDDLPLVAEDPHQ 262
Db 132 GHVMNALPEDAKAWEHGVFVDSADVIRPP-----SASATNDPLEVTVEFSE 181
Qy 263 WYLE-DOFADGAPPLEKAGQVQVGDVTDWEYVKIRMLNAGHVLMCFPGTLVGYENVDDAI 321
Db 182 WIVDKTFKALPNI---PGMEITDNLMAFVERKLTLTNGHTAITAYLGLAGHQIIRAI 239
Qy 322 EDELLGNLKNLNDVIPTLKAPSGMTL-----EGYRDSVTSRFSNKAMSQOT 370
Db 240 LDEKIRAVVKGAMEE-----SGAVLIKRYGFDADKHAAYIQILGRFENPYLKDDV 290
Qy 371 LRIADGCSKVQVFWTETVYRAIEDKRLSRIAFTAGIASVLEMLGRDEKGGTYESEPTY 430
Db 291 ERVG-----RQPLKLSAGDRLIKPLL-----GTLEYGLPHK 322
Qy 431 GDAEWKLAKADDFESSKLPAFDGWRDLDTSELDQKVIIVLRIEKGKAAI 483
Db 323 NLIE-GIAAAMHFR-----SEDDPQAOELALADIADKGPQAL 358

RESULT 2
US-08-928-692-28
; Sequence 28, Application US/08928692
; Patent No 5958727
; GENERAL INFORMATION:
; APPLICANT: Brody, Howard
; APPLICANT: Yaver, Deborah S.
; APPLICANT: Lamsa, Michael
; APPLICANT: Hansen, Kim
; TITLE OF INVENTION: Methods for Modifying the Production of
; TITLE OF INVENTION: a Polypeptide
; NUMBER OF SEQUENCES: 80
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: No. 5958727 of No. 5958727disk of No. 5958727th America, Inc.
; STREET: 405 Lexington Avenue
; CITY: New York
; STATE: NY
; COUNTRY: USA
; ZIP: 10174
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: DOS
; SOFTWARE: FastSEQ for Windows Version 2.0
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/928,692
; FILING DATE: 12-SEPT-1997
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Lambiris, Elias J
; REGISTRATION NUMBER: 33,728
; REFERENCE/DOCKET NUMBER: 4944.200-US
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 212-867-0123
; TELEFAX: 212-878-9655
; INFORMATION FOR SEQ ID NO: 28:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 366 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: No. 5958727e
```

```
US-08-928-692-28
Query Match 7.6%; Score 190; DB 2; Length 366;
Best Local Similarity 22.5%; Pred. No. 4e-10;
Matches 67; Conservative 50; Mismatches 121; Indels 60; Gaps 10;

Qy 29 VHFVGNGFRAHEAFYVEQILEHAPDMAIVGVLGTGSDRSKKKAEFEKQAQCLXSLTETA 88
Db 4 LHFAGNIGRG-----FIGALLHHS-----GYDVF-----A 30
Qy 89 PSCKSTVRVMGALRDYLLAPADPEAVLKLHVDPAIRIVSMITTEGGYININETTGAFDLEN 148
Db 31 DVNETMWSLLNEKKEKTYVELAE-EGRSSEITGPVSAINSGSQTEELYRL-----MNE 81
Qy 149 AAVKADLNKPEKPSVFGVYVEALRRRRDAGGKAFVTMSCDNLRHNGNVARKAPLGYAKA 208
Db 82 AALITAVCPNVKLIIAPSIAGELRRNTA--NTLIIACENMIGGSSFLKKEI--YSHL 137
Qy 209 RDELAKWIEENATEPNGMVDRIPTVSAEIAKLNAAAGLDDDLPLVAEDPHQWVLEDO 268
Db 138 TEAE-QKVSSETLGPNSAVDRIVPIQHHE-----DPLKVSVEPFFEWVIDES 184
Qy 269 FADGRPPLEKAGQVQVGDVTDWEYVKIRMLNAGHVLMCFPGTLVGYENVDDAIDSEL 326
Db 185 GFKGKTPVIN-GALEVDDLTPIERKLTFTVNTGHAHTAYGVQYRGLKTVKKAIDHPEI 241

RESULT 3
US-09-339-972-28
; Sequence 28, Application US/09339972
; Patent No. 6323002
; GENERAL INFORMATION:
; APPLICANT: Brody, Howard
; APPLICANT: Yaver, Deborah S.
; APPLICANT: Lamsa, Michael
; APPLICANT: Hansen, Kim
; TITLE OF INVENTION: Methods for Modifying the Production of
; TITLE OF INVENTION: a Polypeptide
; NUMBER OF SEQUENCES: 80
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: No. 6323002 of No. 6323002disk of No. 6323002th America, Inc.
; STREET: 405 Lexington Avenue
; CITY: New York
; STATE: NY
; COUNTRY: USA
; ZIP: 10174
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: DOS
; SOFTWARE: FastSEQ for Windows Version 2.0
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/339,972
; FILING DATE:
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/928,692
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: Lambiris, Elias J
; REGISTRATION NUMBER: 33,728
; REFERENCE/DOCKET NUMBER: 4944.200-US
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 212-867-0123
; TELEFAX: 212-878-9655
; INFORMATION FOR SEQ ID NO: 28:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 366 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: No. 6323002e
US-09-339-972-28
```

Query Match 6.4%; Score 162; DB 2; Length 391;
Best Local Similarity 21.8%; Pred. No. 2.4e-07;
Matches 77; Conservative 55; Mismatches 149; Indels 72; Gaps 15;

; TOPOLOGY: linear
 ; MOLECULE TYPE: NO. 6323002e
 US-09-339-972-26

368	SS	EOPLRLGALKSNIGHTQAAAAGVAGVKKVQAMRHGLLPKTLHVDPEPSDQIDWSAGTVE	427
183	FTV	MSCD-NLRHNGNVARKAFLGYAKARDP-----ELAKWIEENATF--PNG-----MV	228
428	LL	EAVDPEKQDGLRRAAVSFGISGTNAHVHVEEAPAVEDSPAVEPPAGGCVVPPV	487
229	DRITP-TV	SAIEAKKLNAAAGLDDDLPLVAEDFQHWLEDOQFADGRPPLEKAGVQMVGDV	287
488	SAKTPAALDAQIGOLAAYADGRTDVP	AVA-----ARALYDSTAMEHRAV-AVGDS	538
288	TDWEYV	KIRM-----LNAGHVMLCFPG-----ILVGYENVDDAIEDSELLGNLKNY	3333
539	REALRDALRMPEGLV	RGCTSSDVGEVAFVFPQGQGTQWAGMGAELLDSGPEAASKACETA	598
334	LNKDVIPTLKA-----PSGMTLE-----GYRDSVIRSFNKAMSQTL-----	371	
599	LSRYVDMSLEAVVRQEPGAPT	LDRVDVQVPVTEAVMVSLAKVQHGHGTTPOAVVHGSQGE	658
372	-----RTASDCCKVQVFWTETVR	RAIEDKRLSRIAFGTASVLEMLRGDEKG---	420
659	IAAAVYAGATLDDAARVTL	RSKSIAAHLAGGGMISLALDEAAVLKRLSDFDGSLVAA	718
421	-----GTYESSEPTYDAEWNK	LAKADDFESSLKLPAFDGWRDLDTSLDQKVIVLRKIIIR	475
719	VNGPTATVWSCDPTQIEELARTCEADG	VRARI-IP-----VDYASHSQRVEIIEKELA	770
476	E	476	
771	E	771	

```

RESULT 7
US-09-105-537-35
US-09-105-537-35
Sequence 35, Application US/09105537A
Patent No. 6265202
GENERAL INFORMATION:
APPLICANT: Sherman, D.H.
APPLICANT: Liu, H.
APPLICANT: Xue, Y.
APPLICANT: Zhao, L.
FILE OF INVENTION: DNA encoding methymycin and pikromycin
FILE REFERENCE: 600.438US1
CURRENT APPLICATION NUMBER: US/09/105,537A
CURRENT FILING DATE: 1998-06-26
NUMBER OF SEQ ID NOS: 43
SOFTWARE: FastSEQ for Windows Version 3.0
SEQ ID NO 35
LENGTH: 1562
TYPE: PRT
ORGANISM: Streptomyces venezuelae
US-09-105-537-35

```

[illegible]

Db 539 REALDALRMEGLVRGTSDDVGRVAFVFPQGCTQWAGMGAELLDSSPEFAASMAECETA 598
QY 334 LNKDVIPTLKA-----PSGMTLE-----GYRDSVISRESNKAMSQDTL----- 371
Db 599 LSRVDMVSLVAVVQRPQAPGLDRVDVQVQVTFVAVVSLAKVWQHGGITPQAVVGHSGQE 658
QY 372 -----RIASDGCCKVQVFTETVRRRAIEDKRLDSRIAFTAGTASYLEMLRGRDEKG---- 420
Db 659 IAAAYVAGALTLDAAARVVTLRSKSTAAHLAGKGMISLALDEAAVLKRLSDFDGLSVAA 718
QY 421 -----GTYESSEPTYGDAEWKLAADDFESSLKLPAFDGWRDLDTSELQKQVIVLRKIIR 475
Db 719 VNGPTATVVGSDPTQIEELARTCEADGVRARI-IP-----VDYASHSRQVEIEKELA 770
QY 476 E 476
Db 771 E 771
RESULT 8
US-09-105-537-6
; Sequence 6, Application US/09105537A
; Patent No. 6265202
; GENERAL INFORMATION:
; APPLICANT: Sherman, D.H.
; APPLICANT: Liu, H.
; APPLICANT: Xue, Y.
; APPLICANT: Zhao, L.
; TITLE OF INVENTION: DNA encoding methymycin and pikromycin
; FILE REFERENCE: 600.438US1
; CURRENT APPLICATION NUMBER: US/09/105,537A
; CURRENT FILING DATE: 1998-06-26
; NUMBER OF SEQ ID NOS: 43
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 6
; LENGTH: 11877
; TYPE: PRT
; ORGANISM: Streptomyces venezuelae
US-09-105-537-6

Query Match 4.2%; Score 105.5; DB 4; Length 11877;
Best Local Similarity 20.8%; Pred. No. 20;
Matches 100; Conservative 74; Mismatches 194; Indels 113; Gaps 21;
QY 87 TAPSGKSTVRVM-GALRDYLLAPADPEAVLKH-----LVDPAIRIVSMITTEGGYNINET 140
Db 9001 TAPHGFSQQRVIRRALADARLAPGDVDVVEAHGTGTRLGDPLAQALLIATYG-----QEK 9055
QY 141 TGAFDLENAAVRADLKNPEKSPVFGYV--VEALRR-----RWDAGGKA 182
Db 9056 SSEQPLRLGALKSNIGHTQAAAGVAGVIRKMQAMRHGLLPKTLHVDPEPSDQIDWSAGTVE 9115
QY 183 FTVMSCD-NLRHNGNVARKAFGLYAKARDP-----ELAKWIEENATF--PNG-----NV 228
Db 9116 LITEAVDWEPEKQDGGURRAVSSFGISGTAHVLEAPAVEDSPAVEPPAGGGVVPWPV 9175
QY 229 DRITP-TVSAEIAKLNAAAGSLDDDLPLVAEDFHQWVLEDDQFADGRPPLEKAGVQWVGVD 287
Db 9176 SAKTPAALDAQICOLAAAYADGRDVIDPVA-----ARALVDSSTANEHRAV-AVGDS 9226
QY 288 TWWEYVKIRM-----LNAGHVMCLCPFG-----ILVGYENVDDAIEDSELLGNLKNY 333
Db 9227 REALDALRMEGLVRGTSDDVGRVAFVFPQGCTQWAGMGAELLDSSPEFAASMAECETA 9286
QY 334 LNKDVIPTLKA-----PSGMTLE-----GYRDSVISRESNKAMSQDTL----- 371
Db 9287 LSRVDMVSLVAVVQRPQAPGLDRVDVQVQVTFVAVVSLAKVWQHGGITPQAVVGHSGQE 9346
QY 372 -----RIASDGCCKVQVFTETVRRRAIEDKRLDSRIAFTAGTASYLEMLRGRDEKG---- 420
Db 9347 IAAAYVAGALTLDAAARVVTLRSKSTAAHLAGKGMISLALDEAAVLKRLSDFDGLSVAA 9406

QY 421 -----GTYESSEPTYGDAEWKLAADDFESSLKLPAFDGWRDLDTSELQKQVIVLRKIIR 475
Db 9407 VNGPTATVVGSDPTQIEELARTCEADGVRARI-IP-----VDYASHSRQVEIEKELA 9458
QY 476 E 476
Db 9459 E 9459
RESULT 9
US-08-336-308A-10
; Sequence 10, Application US/08336308A
; Patent No. 6017532
; GENERAL INFORMATION:
; APPLICANT: Travis, James
; APPLICANT: Potempa, Jan S.
; APPLICANT: Barr, Philip J.
; APPLICANT: Pavloff, Nadine
; TITLE OF INVENTION: Porphyromonas gingivalis
; TITLE OF INVENTION: Arginine-specific Proteinase Coding Sequences
; NUMBER OF SEQUENCES: 16
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Greenlee, Winner and Sullivan, P.C.
; STREET: 5370 Manhattan Circle, Suite 201
; CITY: Boulder
; STATE: Colorado
; COUNTRY: US
; ZIP: 80303
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/336,308A
; FILING DATE: 08-NOV-1994
; CLASSIFICATION: 435
; PRIOR APPLICATION NUMBER:
; APPLICATION NUMBER: US 08/119,361
; FILING DATE: 10-SEP-1993
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/265,441
; FILING DATE: 24-JUN-1994
; ATTORNEY/AGENT INFORMATION:
; NAME: Ferber, Donna M.
; REGISTRATION NUMBER: 33,878
; REFERENCE/DOCKET NUMBER: 21-93C
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (303) 499-8080
; TELEFAX: (303) 499-8089
; INFORMATION FOR SEQ ID NO: 10:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1704 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-08-336-308A-10

Query Match 4.2%; Score 105; DB 3; Length 1704;
Best Local Similarity 21.6%; Pred. No. 0.95;
Matches 99; Conservative 60; Mismatches 186; Indels 114; Gaps 22;
QY 86 ETAPSGKSTVRVMGALRDYLLAPADPEAVLKHLDVDPAIRIVSMITTEGGYNINETTGAFD 145
Db 752 QVPSDTHTLWPCNSVPANLFAFF--EYTVPENADPSCSPTNM-IMDGTASVNIPTAGTD 808
QY 146 LENAAYKADLK---NPEKPSVFGYVVEA-----LRREWDAG----- 179
Db 809 FAIAAPQANAKIWIAGQGTKEDDYVEAGKKVHFLMKMGSGDGTFLTISEGGSDITY 868
QY 180 -----CKAFTVMSCDNLRHNGNVARKAFGLYAKARDPELAKVTEENATFPNGV 228
Db 869 TVYRGTIKI EGLTATTFEEDGVA-TGNHEYCVKRYTAGVSPKVC KDVTVEGSNEFAPV 927


```

; INFORMATION FOR SEQ ID NO: 10:
;
; SEQUENCE CHARACTERISTICS:
;   LENGTH: 1704 amino acids
;   TYPE: amino acid
;   TOPOLOGY: linear
;
; MOLECULE TYPE: protein
US-09-490-931-10

```

Query Match 4.2%; Score 105; DB 4; Length 1704;
Best Local Similarity 21.6%; Pred. No. 0.95;
Matches 99; Conservative 60; Mismatches 186; Indels 114; Gaps 22;

QY	86	ETAPSGKSTVRVNGALRDYLLAPADAEAVLKLHVPATRIVSMITTECGYVNIINTEGAFD	145
Db	752	QVTPSDTHLWPNCSVPANLFAPE---EXTVPENADPSCSPNTNM-IMDGTSVNIIPAGTYD	808
QY	146	LENAAVKADLK---NPEKEPSTVFVGWVEA-----LRRRWDAAG-----	179

Db 809 FAIAPOANAKIWIAGCQPTKEDDYVFEAGKKYHFLMKMGSGDGTELTISEGGSDYTY 868

QY 180 -----GKFTVMSCDNLRHNGNVARKAFILGYAKARDPELAKWIEENATFPNGMV 228

DB 869 TVYRDGFIKKEGLTATITFEEDGVA-TGNHEXCVVEVKYTAGVSPRYCKDVTVEGSNEFAPV 927

[illegible]

QV 288 TDWEYVKIRMLNAGHVMJCFPGILVGYENVDDATEDSEJIGNI-----KNVINKDVIPTI. 342

RESULT 12
PCT-US95-13749-4
; Sequence 4, Application PC/TUS9513749
; GENERAL INFORMATION:
; APPLICANT: Amgen Inc.
; TITLE OF INVENTION: COMPOSITIONS FOR INCREASED
; BIOAVAILABILITY OF ORALLY DELIVERED THERAPEUTIC AGENTS

```

; STRANDEDNESS: unknown
; TOPOLOGY: unknown
; MOLECULE TYPE: protein
PCT-US95-13749-4

```

Query Match 4.1%; Score 104; DB 5; Length 588;
Best Local Similarity 19.9%; Pred. No. 0.21;
Matches 105; Conservative 66; Mismatches 176; Indels 180; Gaps 29;

Qy 17 PPYDIDGIPGIVHFGVGNFFRAHEAF--YVEQ--ILEHAPDWAIVGVLTGSDRSKKA 72
Db 74 PQVAATGDPDII-----FWAHDFGGYAGSGLLAEITPDKAF----- 111

Qy 73 EEFKAQDCLYSLTETAPSGKSTVRVNGALRDYLLA-----PADPE----- 112

Db 112 -----QDKLYPFTWDA-----VRYNGKLIAYPIAVEALSILYNKDLLPNPPKTWEETPA 160

QY -----AVLAHLVDPAIRKIVSWIIIEGGYNINETIGAFDL-----ENAAVRADL- 139
113 |::|:| : ||| |::| ::||

Qv 156 -----KNDKDPSTVECVIVFAI PPPNDACCCKA ETV -----MSCONI PUNCNVA ----- 108

Db 220 FLVDLIKKNHNMADTDYSIAEA---AFNKGETAMTINGPWAWSNIDTSKVNYGTVLPTF 276

QY 199 ---RKAFGLYAKARDEPELAKWIEENATFPNGMVDRIPTVSAEIAKKL-----NAA 246

Db 277 KGQPSKPFVGVLSA-----GINAASP-----KELAKEFLENYLLTDEGL 316

```

RESULT 13
US-08-955-957A-2
; Sequence 2, Application US/08955957A
; Patent No. 6312920
; GENERAL INFORMATION:
; APPLICANT: Dehoff, Bradley S.
; APPLICANT: Rosteck Jr., Paul R.
; TITLE OF INVENTION: SAM Operon
; NUMBER OF SEQUENCES: 6
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Eli Lilly and Company
; STREET: Lilly Corporate Center
; CITY: Indianapolis
; STATE: Indiana
; COUNTRY: U.S.
; ZIP: 46285
; COMPUTER READABLE FORM: disk
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/955,957A
; FILING DATE:
; CLASSIFICATION:
; ATTORNEY/AGENT INFORMATION:
; NAME: Webster, Thomas.D.

```


TITLE OF INVENTION: PREDISPOSITION TO PEPTIC ULCERATION AND GASTRIC CARCINOMA
NUMBER OF SEQUENCES: 4
CORRESPONDENCE ADDRESS:
ADDRESSEE: NEEDLE & ROSENBERG, P.C.
STREET: 127 Peachtree Street, Suite 1200
CITY: Atlanta
STATE: Georgia
COUNTRY: USA
ZIP: 30303
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/316,397B
FILING DATE:
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: SPRATT, GWENDOLYN D.
REGISTRATION NUMBER: 36,016
REFERENCE/DOCKET NUMBER: 2200.030
TELECOMMUNICATION INFORMATION:
TELEPHONE: 404/688-0770
TELEFAX: 404/688-9880
INFORMATION FOR SEQ ID NO: 2:
SEQUENCE CHARACTERISTICS:
LENGTH: 859 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-316-397B-2

Query Match 4.0%; Score 101; DB 1; Length 859;
Best Local Similarity 18.8%; Pred.No. 0.76;
Matches 103; Conservative 84; Mismatches 196; Indels 166; Gaps 25;
QY 22 DGIKP---GIVHFGVGNFPRAHE-----AFYVEQILEHAPDWAIVG---- 59
DB 308 DGIEPEKVSLLYAGNGFGGAKHDWATVGYKQDQGNVATIINVHMKNGSLVIAGGEGK 367
QY 60 -----VGLTGSDRSKKAE-----EKAOD--CLYSITETAPSGKSTVRVM 98
DB 368 INNPFFLYKEDQLGQSQRALSOEQIKNIDMEFLAQNNAKLDLSLSE-----KEKEFR 422
QY 99 GALRDYLLAPADPEAVLKHLDVPAIRIVSMITTEGGYNINETTGAFLDLENAAVKADLKNP 158
DB 423 NEIKDF---QKDSKPYLDALGNDRIFAIVSKPKPKHSALITEF-----KGDLSYT 469
QY 159 EKPSTVFG-YVVEALRRRWDAAGKAFVMSCDNLRHNGVARKAFGLYAKARDPKELAKWI 217
DB 470 LK---VMGKKQIKALDRE-----KNVTLOG--NLKHDGVN----FVNSNFKYTNASKSP 515
QY 218 EENATFPNGMVD-----RITVSAEIAKKNLAASGLDDDLPLVAE 258
DB 516 NKGVGVTNGVSHLEAGFSKAVFNPINLPNLNLAITSVRRDLEDKLIAGLSPOEANKLVK 575
QY 259 DFHQWLEDFQADGRPPLEKAGVQVGDVTDWEYVKIRMLNAGHVMLCPFGILVGVENVYD 318
DB 576 DF-----LSSNKLGVKALNFENKAVAEAKNTGN-----YDEVK 508
QY 319 DAIEDSELGNLKNYLNKNDIPTLKAPSG-----MTLEGYRDSVISRFSNKAMSD-Q 369
DB 609 RAQKDLKSLKKREHLEKDVAKNLESKSGKNKNKMEAKQAQNSQKDEIFALINKNRRDAR 668
QY 370 TLRIASDGCCKVQVFWETVVRRAIEDKRLSRIAGIASYLEMLRG-RDEKGGTYESSEP 428
DB 669 ATAYAQN-----LKGIKRELSDK--LENINKDLKDFSKSFDGFKNGKNKDFSKAE 717
QY 429 TYG-----DAEWKLAKADDFESSKLKLPAPDQWRDLDTSELDQ-----KVI 468
DB 718 TLKALKGSKVDLGINPEW-ISKVENLNAALN--EFKNGKNKDFSKVTQAKSQENSIKDV 774

QY 469 VLRKIIREK 477
DB 775 IINQKITDK 783
Search completed: May 8, 2003, 16:25:50
Job time : 42 secs

THIS PAGE BLANK (USPTO)

GenCore version 5.1.4_p5_4578
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: May 8, 2003, 16:24:41 ; Search time 53 Seconds
(without alignments)
842.122 Million cell updates/sec

Title: US-09-926-163B-2

Perfect score: 2516

Sequence: 1 MITRETLKSLPANVOAPDY.....KVIVLRTIREKGVKAIPA 485

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 349150 seqs, 92025710 residues

Total number of hits satisfying chosen parameters: 349150

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Published Applications_AA:*

- 1: /cgn2_6/ptodata/1/pubpaa/US08_NEW_PUB.pep.*
- 2: /cgn2_6/ptodata/1/pubpaa/PCT_NEW_PUB.pep.*
- 3: /cgn2_6/ptodata/1/pubpaa/US06_NEW_PUB.pep.*
- 4: /cgn2_6/ptodata/1/pubpaa/US06_PUBCOMB.pep.*
- 5: /cgn2_6/ptodata/1/pubpaa/US07_NEW_PUB.pep.*
- 6: /cgn2_6/ptodata/1/pubpaa/US07_PUBCOMB.pep.*
- 7: /cgn2_6/ptodata/1/pubpaa/PCTUS_PUBCOMB.pep.*
- 8: /cgn2_6/ptodata/1/pubpaa/US08_PUBCOMB.pep.*
- 9: /cgn2_6/ptodata/1/pubpaa/US09_NEW_PUB.pep.*
- 10: /cgn2_6/ptodata/1/pubpaa/US09_PUBCOMB.pep.*
- 11: /cgn2_6/ptodata/1/pubpaa/US10_NEW_PUB.pep.*
- 12: /cgn2_6/ptodata/1/pubpaa/US10_PUBCOMB.pep.*
- 13: /cgn2_6/ptodata/1/pubpaa/US60_NEW_PUB.pep.*
- 14: /cgn2_6/ptodata/1/pubpaa/US60_PUBCOMB.pep.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	802	31.9	503	9	US-09-738-626-3626
2	221.5	8.8	384	10	US-09-815-242-10595
3	205	8.1	382	10	US-09-815-242-10385
4	200	7.9	378	10	US-09-815-242-13688
5	181	7.2	368	10	US-09-815-242-5636
6	181	7.2	368	10	US-09-815-242-12393
7	125	5.0	1504	9	US-09-932-145-7
8	107	4.3	1703	9	US-09-824-574-3
9	107	4.3	1703	10	US-09-801-368-340
10	105.5	4.2	1562	9	US-09-860-846-35
11	105.5	4.2	1562	9	US-09-988-384B-35
12	105.5	4.2	1562	10	US-09-861-289-35
13	105.5	4.2	11877	9	US-09-860-846-6
14	105.5	4.2	11877	10	US-09-861-289-6
15	105.5	4.2	12199	9	US-09-988-384B-6
16	103	4.1	669	9	US-10-083-357-1336
17	102	4.1	407	9	US-09-943-702-2
18	100.5	4.0	1338	10	US-09-402-100-4
19	100	4.0	762	10	US-09-815-242-11058

20	99.5	4.0	2910	9	US-10-124-800-2
21	98	3.9	559	10	US-09-815-242-11904
22	98	3.9	852	9	US-09-738-626-6999
23	97.5	3.9	736	9	US-09-978-295A-526
24	97.5	3.9	736	9	US-09-978-697-526
25	97.5	3.9	736	9	US-09-978-192A-526
26	97.5	3.9	736	9	US-09-998-832A-526
27	97.5	3.9	736	9	US-09-978-189-526
28	97.5	3.9	736	9	US-10-174-590-420
29	97.5	3.9	736	9	US-10-176-758-420
30	97.5	3.9	736	9	US-10-175-737-420
31	97.5	3.9	736	9	US-10-173-706-420
32	97.5	3.9	736	9	US-10-175-738-420
33	97.5	3.9	736	9	US-10-175-752-420
34	97.5	3.9	736	9	US-10-176-482-420
35	97.5	3.9	736	9	US-10-176-757-420
36	97.5	3.9	736	9	US-10-176-913-420
37	97.5	3.9	736	9	US-10-180-552-420
38	97.5	3.9	736	9	US-10-180-557-420
39	97.5	3.9	736	9	US-10-173-700-420
40	97.5	3.9	736	9	US-10-174-572-420
41	97.5	3.9	736	9	US-10-174-579-420
42	97.5	3.9	736	9	US-10-174-582-420
43	97.5	3.9	736	9	US-10-174-588-420
44	97.5	3.9	736	9	US-10-175-739-420
45	97.5	3.9	736	9	US-10-175-740-420

ALIGNMENTS

RESULT 1

US-09-738-626-3626
; Sequence 3626, Application US/09738626
; Publication No. US20020197605A1
; GENERAL INFORMATION:
; APPLICANT: NAKAGAWA, SATOSHI
; APPLICANT: MIZOGUCHI, HIROSHI
; APPLICANT: ANDO, SEIKO
; APPLICANT: HAYASHI, MIKIRO
; APPLICANT: OCHIAI, KEIKO
; APPLICANT: YOKOI, HARUHIKO
; APPLICANT: TATEISHI, NAOKO
; APPLICANT: SENO, AKIHIRO
; APPLICANT: IKEDA, MASATO
; APPLICANT: OKAZAKI, AKIO
; TITLE OF INVENTION: NOVEL POLYNUCLEOTIDES
; FILE REFERENCE: 249-125
; CURRENT APPLICATION NUMBER: US/09/738,626
; CURRENT FILING DATE: 2000-12-18
; PRIOR APPLICATION NUMBER: JP 99/377484
; PRIOR FILING DATE: 1999-12-16
; PRIOR APPLICATION NUMBER: JP 00/159162
; PRIOR FILING DATE: 2000-04-07
; PRIOR APPLICATION NUMBER: JP 00/280988
; PRIOR FILING DATE: 2000-08-03
; NUMBER OF SEQ ID NOS: 7059
; SOFTWARE: PatentIn ver. 3.0
; SEQ ID NO 3626
; LENGTH: 503
; TYPE: PRT
; ORGANISM: Corynebacterium glutamicum
US-09-738-626-3626

Query Match 31.9%; Score 802; DB 9; Length 503;

Best Local Similarity 39.0%; Pred. No. 7.4e-63;

Matches 168; Conservative 82; Mismatches 163; Indels 18; Gaps 6;

QY 2 ITRETLKSL--PANVOAPPYDIDIKPGVHGFNFRAHEAFVVEQILEH--APDWAI 57

Db 7 LNTENLQELASTSGVQIFAFNADVPAGVHGFVGVGFRAHQAWYLNELMNECKALDWCI 66

QY 58 VGVGLTGSRSRKKAEFPKQDCLYSLTETAPSGKSTVRVMGALRDYLLAPADPEAVLKH 117

Db 67 ICGVMPDSVRMRDA--LASQDHLTYLTAKPDGTLQDQKIIGSIIDYVFPAPEDPARAVAT 124
Qy 118 LVDPATRIYSMTITTEGYNINTEGAFDLENAAVKADLK-----NPEKSTVFGVYVVEAL 172
Db 125 LAQDSIRIVSLVTEGGYNIDPATEDFDITNPRIVADREALQAGDTSTLQTFEGLITAA 184
Qy 173 RRRWDAGGKRAFTVMSCDNLNRHNGNVARKAFGLYAKARDPELAKWIEENATFPNGMVDRT 232
Db 185 ISKESGSTPTIMSDNIQNGDLAKRFLAFASHVSSELGEWENNVAFPNSMVDRT 244
Qy 233 PTVS---ABIANKLNASGLDDDLPLVADDFHQLVLEDOFADGRPLEKAGVQMGVDYTD 289
Db 245 PETTDGDRDIKEI---GYIDAMPVVSDFQWLEDAFTQRPAYEEVGVVSDVEP 300
Qy 290 WEVVKRLNAGHVMLCPGILVGVENVDDAJEDELGNLKNLKNKDVIPTLKAPSGMT 349
Db 301 YELMKRLNLNASHQGLCYFGLHGHVHVDWADRFQDFLLAYWERATPTLKELPGVD 360
Qy 350 LEGYRDSVISRFSNKMAMSDQTLRIASDGCSCVQVFTWETVTRRAIEDKRLSLRIAFGIASY 409
Db 361 LDAYRQRLIARFGNAVKDTPRLCAESSDRIPKLLPVVRENLAAGRDVTLSSAIVASW 420
Qy 410 LEMLRGRDEK 420
Db 421 ARYAGCTDEQ 431

RESULT 2
US-09-815-242-10595
; Sequence 10595, Application US/09815242
; Patent No. US20020061569A1
; GENERAL INFORMATION:
; APPLICANT: Haselbeck, Robert
; APPLICANT: Ohlsen, Kari L.
; APPLICANT: Zyskind, Judith W.
; APPLICANT: Wall, Daniel
; APPLICANT: Trawick, John D.
; APPLICANT: Carr, Grant J.
; APPLICANT: Yamamoto, Robert T.
; APPLICANT: Xu, H. Howard
; TITLE OF INVENTION: Identification of Essential Genes in
; FILE REFERENCE: ELITRA.011A
; CURRENT APPLICATION NUMBER: US/09/815,242
; PRIOR FILING DATE: 2001-03-21
; PRIOR APPLICATION NUMBER: 60/191,078
; PRIOR FILING DATE: 2000-03-21
; PRIOR APPLICATION NUMBER: 60/206,848
; PRIOR FILING DATE: 2000-05-23
; PRIOR APPLICATION NUMBER: 60/207,727
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: 60/242,578
; PRIOR FILING DATE: 2000-10-23
; PRIOR APPLICATION NUMBER: 60/253,625
; PRIOR FILING DATE: 2000-11-27
; PRIOR APPLICATION NUMBER: 60/257,931
; PRIOR FILING DATE: 2000-12-22
; PRIOR APPLICATION NUMBER: 60/269,308
; PRIOR FILING DATE: 2001-02-16
; NUMBER OF SEQ ID NOS: 14110
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 10595
; LENGTH: 384
; TYPE: PRT
; ORGANISM: Enterococcus faecalis
US-09-815-242-10595

Query Match 8.8%; Score 221.5; DB 10; Length 384;
Best Local Similarity 22.6%; Pred. No. 1.9e-11;
Matches 83; Conservative 61; Mismatches 123; Indels 101; Gaps 14;

Qy 29 VIFGVGNFRAEAFVVEQILEHAPDWAIVGVLGTSDRSKKAEFFKAQDCLYSLTETA 88

Db 4 VHFAGAGTGRG---FIGEIL-----AKNGFHITFVD-----VNET- 35
Qy 89 PSGKSTVRVMGALRDYLLAPADPEAVLKHLDPAIRIVSMITTEGYNINTEGAFDLEN 148
Db 36 -----IIQALKERKSYTTELADAS-----HQQINVENVT-----GLNNWTEPEKVVE 77
Qy 149 AAVKADLNKPEKSTVFGVYVVEALRRWDAGKA-----FTVMSCDNLNRHNGNVARKAPL 203
Db 78 ATAEDADVTTAIGPNILPRIAELIAQIGIDARAEANCQKPLDIIACENM-----IGGSTFL 132
Qy 204 GYAKARDELAKWIEENA-----TFPNGMVDRTPTVSAAETIAKKLNASGLDDDLPLVA 257
Db 133 A-----EVAKYLNKPAYAEQWIGFPPDAADVRIVPLOKHE-----DFLFGVQV 174
Qy 258 EDFHQLVLEDOFADGRPLEKAGVQMGVDYTDWEVVKRLNAGHVMLCFPGILVGYENV 317
Db 175 EPCFCEWIDDTNRKAK-ELQLEGVHYVADLEPIYERKLFVSNTGHTATVAYTGALLGYQTI 233
Qy 318 DDAIEDSELLGNLKNLKNKDVIPTLKAPSGMTL-----EGYRDSVISRFSNKMAM 366
Db 234 DEAMQDALVVAQLKSVLOE-----TGKLLVAKWNFDEQEAHAAYIEKIIORFONKYI 284
Qy 367 SDQTLRIA 374
Db 285 SDAITRVA 292

RESULT 3
US-09-815-242-10385
; Sequence 10385, Application US/09815242
; Patent No. US20020061569A1
; GENERAL INFORMATION:
; APPLICANT: Haselbeck, Robert
; APPLICANT: Ohlsen, Kari L.
; APPLICANT: Zyskind, Judith W.
; APPLICANT: Wall, Daniel
; APPLICANT: Trawick, John D.
; APPLICANT: Carr, Grant J.
; APPLICANT: Yamamoto, Robert T.
; APPLICANT: Xu, H. Howard
; TITLE OF INVENTION: Identification of Essential Genes in
; FILE REFERENCE: ELITRA.011A
; CURRENT APPLICATION NUMBER: US/09/815,242
; PRIOR FILING DATE: 2001-03-21
; PRIOR APPLICATION NUMBER: 60/191,078
; PRIOR FILING DATE: 2000-03-21
; PRIOR APPLICATION NUMBER: 60/206,848
; PRIOR FILING DATE: 2000-05-23
; PRIOR APPLICATION NUMBER: 60/207,727
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: 60/242,578
; PRIOR FILING DATE: 2000-10-23
; PRIOR APPLICATION NUMBER: 60/253,625
; PRIOR FILING DATE: 2000-11-27
; PRIOR APPLICATION NUMBER: 60/257,931
; PRIOR FILING DATE: 2000-12-22
; PRIOR APPLICATION NUMBER: 60/269,308
; PRIOR FILING DATE: 2001-02-16
; NUMBER OF SEQ ID NOS: 14110
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 10385
; LENGTH: 382
; TYPE: PRT
; ORGANISM: Escherichia coli
US-09-815-242-10385

Query Match 8.1%; Score 205; DB 10; Length 382;
Best Local Similarity 21.9%; Pred. No. 5.4e-10;
Matches 104; Conservative 71; Mismatches 161; Indels 138; Gaps 20;

Qy 29 VIFGVGNFRAEAFVVEQILEHAPDWAIVGVLGTSDRSKKAEFFKAQDC--LYSLTE 86

Db 4 LHFGAGNIGRG-----FIGILLADA-----GIQLTADVQVVLDAALNARHSQVHVWGE 53
QY 87 T-----APSGKSTVRVMGALRDYLLAPADPEAVLKHVLVDPAIRIVSMITTEGGYNINETTG 142
Db 54 TEQVDTVSGVNAVSSIGDDVDVJIAQVD-----LVTTAV-----G 88
QY 143 AFLENAAYKADLNKPEKSTVFYVVEALRRRDWAGGKAFVMSCDNLRHNGNVARKAF 202
Db 89 PVYLERIA-----PAIAKGQV-----KRKEOGNESPLNIIACENM-VRGTTOLK-- 131
QY 203 LGYAKARDPELAK-WIEENATFNGMVDRIPTVSAEIAKLNAASGLDDDLPLVAEDPH 261
Db 132 -GHVMNALPEDAKAWYEEHVFVDSAVDRIVPP-----SASATNDPLEVTVETES 180
QY 262 OWYLE-DQFADGPPLEKAGVQMGVDVTWEYVKIRMLNAGHVMLCFPGILVGYENVDDA 320
Db 181 EWIVDTQFGALPNI--FGMELTDNLMAFVERKLTFTLTGHAITAYLKLAGHQTIROA 238
QY 321 IEDELLGNLKNYLNKDVIPTLKAPSGMTL-----EGYRDSVISRFSNKAMSDQ 369
Db 239 ILDEKIRAVVKGAMEE-----SGAVLIKRYGFDADKHAAYIQILGRFENPYLKDD 289
QY 370 TLRASDGGSKVQVFTETVRRRAIEDKRLSRTAFGIAASYLEMLGRDEKGGTYESSEPT 429
Db 290 VERVG-----RQPLRLKSAGDRLIKPLL-----GTLEYGLPH 321
QY 430 YGDAEWKLAKADDFESSLKLPAFGDWRDLTSELQKVIIVLRKIIREKGVKAAI 483
Db 322 KNLIE-GIAAMHFR-----SEDDPQAOELAAIADKGPQAAI 358

RESULT 4
US-09-815-242-13688
; Sequence 13688, Application US/09815242
; Patent No. US20020061569A1
; GENERAL INFORMATION:

; APPLICANT: Haselbeck, Robert
; APPLICANT: Ohlsen, Kari L.
; APPLICANT: Zyskind, Judith W.
; APPLICANT: Wall, Daniel
; APPLICANT: Trawick, John D.
; APPLICANT: Carr, Grant J.
; APPLICANT: Yamamoto, Robert T.
; APPLICANT: Xu, H. Howard
; TITLE OF INVENTION: Identification of Essential Genes in
; FILE REFERENCE: ELITRA.011A
; CURRENT APPLICATION NUMBER: US/09/815,242
; CURRENT FILING DATE: 2001-03-21
; PRIOR APPLICATION NUMBER: 60/191,078
; PRIOR FILING DATE: 2000-03-21
; PRIOR APPLICATION NUMBER: 60/206,848
; PRIOR FILING DATE: 2000-05-23
; PRIOR APPLICATION NUMBER: 60/207,727
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: 60/242,578
; PRIOR FILING DATE: 2000-10-23
; PRIOR APPLICATION NUMBER: 60/253,625
; PRIOR FILING DATE: 2000-11-27
; PRIOR APPLICATION NUMBER: 60/257,931
; PRIOR FILING DATE: 2000-12-22
; PRIOR APPLICATION NUMBER: 60/269,308
; PRIOR FILING DATE: 2001-02-16
; NUMBER OF SEQ ID NOS: 14110
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 13688
; LENGTH: 378
; TYPE: PRT
; ORGANISM: Streptococcus pneumoniae

Query Match 7.9%; Score 200; DB 10; Length 378;

Best Local Similarity 22.6%; Pred. No. 1.5e-09;
Matches 106; Conservative 71; Mismatches 163; Indels 130; Gaps 21;
QY 29 VHFQVGNFTRAHEAFVVEQIL-----EHPADWAIVGVGLTGSDBSKKKAEEFKAQDCLYSL 84
Db 5 VHFQVGNFTRAHEAFVVEQIL-----EHPADWAIVGVGLTGSDBSKKKAEEFKAQDCLYSL 84
QY 85 TETAPSGKSTVRVMGALRDYLLAPADPEAVLKHVLVDPAIRIVSMITTEGGYNINETTGAF 144
Db 49 IELAQKQOSRIEVTNAG--INSKEHQEVIE-----AIQKTDIITTAIGPNI----- 94
QY 145 DLENAAYKADLNKPEKSTVFYVVEAL-----RRWDAGGKAFVMSCDNLRHNGNVAR 199
Db 95 -----LPFIAELLAKGTEARRVAGNTQALDVMACENM-----IGG 129
QY 200 KAFGLYAKAR---DPELAKWIEENATFNGMVDRIPTVSAEIAKLNAASGLDDDLPLV 256
Db 130 SQFL-YQEVKKYLSPEGLTFADNYIGFPNAAVDRIVPTQSH-----DSLFVM 176
QY 257 AEDFHQWLEDOFADGPPLEKAGVQMGVDVTWEYVKIRMLNAGHVMLCFPGILVGYEN 316
Db 177 VEPFNEWVETKRLK-NPDLRLDEDVHYEEDLEPFFIERKLFVNSGHATSAYIGAHHYGAKT 235
QY 317 VDDATEDSELGNLKNYLNKDVIPTLKAP---SGMPTLEGYRDSVISRFSNKAMSDQTLRI 373
Db 236 ILEALQNPNIKSRIESVL-AEIRSLLIAKWNFDKKELENYHKVIIERFENPFIVDEVSrv 294
QY 374 ASDGSKVQVFTETVRRRAIEDKRLSRTAFGIAASYLEMLGRDEKGGTYESSEPTYGDA 433
Db 295 ARTPIRKIG--YNERFIRPIRELKLS-----LSYKNLLK-----TVG--- 330
QY 434 EWKLAKADDFESSLKLPAFGDWRDLTSELQKVIIVLRKIIREKGVKAAI 483
Db 331 -----YAD-YRDVN-----DEESIRLGELLAKQSVKDVV 359

RESULT 5
US-09-815-242-5636
; Sequence 5636, Application US/09815242
; Patent No. US20020061569A1
; GENERAL INFORMATION:

; APPLICANT: Haselbeck, Robert
; APPLICANT: Ohlsen, Kari L.
; APPLICANT: Zyskind, Judith W.
; APPLICANT: Wall, Daniel
; APPLICANT: Trawick, John D.
; APPLICANT: Carr, Grant J.
; APPLICANT: Yamamoto, Robert T.
; APPLICANT: Xu, H. Howard
; TITLE OF INVENTION: Identification of Essential Genes in
; FILE REFERENCE: ELITRA.011A
; CURRENT APPLICATION NUMBER: US/09/815,242
; CURRENT FILING DATE: 2001-03-21
; PRIOR APPLICATION NUMBER: 60/191,078
; PRIOR FILING DATE: 2000-03-21
; PRIOR APPLICATION NUMBER: 60/206,848
; PRIOR FILING DATE: 2000-05-23
; PRIOR APPLICATION NUMBER: 60/207,727
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: 60/242,578
; PRIOR FILING DATE: 2000-10-23
; PRIOR APPLICATION NUMBER: 60/253,625
; PRIOR FILING DATE: 2000-11-27
; PRIOR APPLICATION NUMBER: 60/257,931
; PRIOR FILING DATE: 2000-12-22
; PRIOR APPLICATION NUMBER: 60/269,308
; PRIOR FILING DATE: 2001-02-16
; NUMBER OF SEQ ID NOS: 14110
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 5636
; LENGTH: 368
; TYPE: PRT

; ORGANISM: Staphylococcus aureus
US-09-815-242-5636

Query Match 7.2%; Score 181; DB 10; Length 368;
Best Local Similarity 21.4%; Pred. No. 7e-08;

Matches 99; Conservative 64; Mismatches 168; Indels 132; Gaps 17;

```
QY 29 VHFVGNGFFRAHEAFYVEOILEHAPDWAIVGVLGTSRKSRAEFAKQDCLYSLET88
    |||||  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
Db 4 VHEGAGNIGRG---FIGYIL-----ADNNVKVTFADVNEELINAL-AHDHGYDVI-LA 51

QY 89 PSCKSTVRVMGALROYLLAPADPEAVLKHLDPAIRIVSMTITEGGYNINETTGAFDLEN 148
    |||||  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
Db 52 DESKTTTRV-----NNVDA-----INSMQPSSEALKQ 77

QY 149 AAVKADLNKPEKPSVFGYVVEALRRRDAGGKFTVMSCDNLHNGNVARKAFGLCYAKA 208
    |||||  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
Db 78 AILEADIITAVGVNLPILIAKSFAPFLKEKTNHVNIVACENAIMATDPLKRAVL----- 132

QY 209 RDELAKTEENATFPNGMVDRTPTVSAEIAKLNAAAGLDDDLPLVAEDFHQWLEDO 268
    |||||  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
Db 133 ---DITGPLGHNHIFANSADVRIPLQKNE-----NILDVMVEPFYEWVVEKD 177

QY 269 FADGRPPLKAGVOMGVDTWDEYVKIRMLNAGHVMCLCPGILGVYENVDDAIEDSELG 328
    |||||  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
Db 178 AWYG-PELNH--IKYVDDLTPTYIERKLLTVNTGHAYLAGKFAKATVLDKAVKSSIEA 234

QY 329 NLKNYLKNDVIPTLKAPSGMTLE-----GYRDSVISRESNKAUSDOTLRIASDGC 379
    |||||  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
Db 235 GLRRVL-----AETSOYITNEFDTEAEQAGYVEKIIDRENNSYLSDEVTRVGRG--- 284

QY 380 KVOVFWTETVRAIEDKRLSIAFGIASYLEMLRGD-EKGGTYESSSEPTYGDAEWKLA 438
    |||||  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
Db 285 -----TLRKIGPKDR-----IKPLKLYLNKDLERTGLLNTAA-----LLL 320

QY 439 KADDFESSLKLPAFDGWRDLDTSELDQKVIIVLRKIIREKGVKA 481
    |||||  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
Db 321 KYDD-----TADQETVEKNNYIKEHGLKA 344
```

RESULT 6

US-09-815-242-12393
Sequence 12393, Application US/09815242

Patent No. US20020061569A1

GENERAL INFORMATION:

APPLICANT: Haselbeck, Robert

APPLICANT: Ohlsen, Kari L.

APPLICANT: Zyskind, Judith W.

APPLICANT: Wall, Daniel

APPLICANT: Trawick, John D.

APPLICANT: Carr, Grant J.

APPLICANT: Yamamoto, Robert T.

APPLICANT: Xu, H. Howard

TITLE OF INVENTION: Identification of Essential Genes in

Prokaryotes

FILE REFERENCE: ELITRA.011A

CURRENT APPLICATION NUMBER: US/09/815,242

CURRENT FILING DATE: 2001-03-21

PRIOR APPLICATION NUMBER: 60/191,078

PRIOR FILING DATE: 2000-03-21

PRIOR APPLICATION NUMBER: 60/206,848

PRIOR FILING DATE: 2000-05-23

PRIOR APPLICATION NUMBER: 60/207,727

PRIOR FILING DATE: 2000-05-26

PRIOR APPLICATION NUMBER: 60/242,578

PRIOR FILING DATE: 2000-10-23

PRIOR APPLICATION NUMBER: 60/253,625

PRIOR FILING DATE: 2000-11-27

PRIOR APPLICATION NUMBER: 60/257,931

PRIOR FILING DATE: 2000-12-22

PRIOR APPLICATION NUMBER: 60/269,308

PRIOR FILING DATE: 2001-02-16

NUMBER OF SEQ ID NOS: 14110

; SOFTWARE: FastSeq for Windows Version 4.0

; SEQ ID NO 12393

; LENGTH: 368

; TYPE: PRT

; ORGANISM: Staphylococcus aureus

US-09-815-242-12393

Query Match

Best Local Similarity 21.4%; Score 181; DB 10; Length 368;

Matches 99; Conservative 64; Mismatches 168; Indels 132; Gaps 17;

```
QY 29 VHFVGNGFFRAHEAFYVEOILEHAPDWAIVGVLGTSRKSRAEFAKQDCLYSLET88
    |||||  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
Db 4 VHEGAGNIGRG---FIGYIL-----ADNNVKVTFADVNEELINAL-AHDHGYDVI-LA 51

QY 89 PSCKSTVRVMGALROYLLAPADPEAVLKHLDPAIRIVSMTITEGGYNINETTGAFDLEN 148
    |||||  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
Db 52 DESKTTTRV-----NNVDA-----INSMQPSSEALKQ 77

QY 149 AAVKADLNKPEKPSVFGYVVEALRRRDAGGKFTVMSCDNLHNGNVARKAFGLCYAKA 208
    |||||  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
Db 78 AILEADIITAVGVNLPILIAKSFAPFLKEKTNHVNIVACENAIMATDPLKRAVL----- 132

QY 209 RDELAKTEENATFPNGMVDRTPTVSAEIAKLNAAAGLDDDLPLVAEDFHQWLEDO 268
    |||||  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
Db 133 ---DITGPLGHNHIFANSADVRIPLQKNE-----NILDVMVEPFYEWVVEKD 177

QY 269 FADGRPPLKAGVOMGVDTWDEYVKIRMLNAGHVMCLCPGILGVYENVDDAIEDSELG 328
    |||||  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
Db 178 AWYG-PELNH--IKYVDDLTPTYIERKLLTVNTGHAYLAGKFAKATVLDKAVKSSIEA 234

QY 329 NLKNYLKNDVIPTLKAPSGMTLE-----GYRDSVISRESNKAUSDOTLRIASDGC 379
    |||||  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
Db 235 GLRRVL-----AETSOYITNEFDTEAEQAGYVEKIIDRENNSYLSDEVTRVGRG--- 284

QY 380 KVOVFWTETVRAIEDKRLSIAFGIASYLEMLRGD-EKGGTYESSSEPTYGDAEWKLA 438
    |||||  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
Db 285 -----TLRKIGPKDR-----IKPLKLYLNKDLERTGLLNTAA-----LLL 320

QY 439 KADDFESSLKLPAFDGWRDLDTSELDQKVIIVLRKIIREKGVKA 481
    |||||  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
Db 321 KYDD-----TADQETVEKNNYIKEHGLKA 344
```

RESULT 7

US-09-932-145-7

Sequence 7, Application US/09932145

Patent No. US20020161191A1

GENERAL INFORMATION:

APPLICANT: Feder, John N.

APPLICANT: Mintier, Gabe

APPLICANT: Kinney, Gene G

APPLICANT: Ramanathan, Chandra S

TITLE OF INVENTION: NOVEL IMIDAZOLINE RECEPTOR HOMOLOGS

FILE REFERENCE: D0020 NP

CURRENT APPLICATION NUMBER: US/09/932,145

CURRENT FILING DATE: 2001-08-17

NUMBER OF SEQ ID NOS: 11

SOFTWARE: PatentIn Ver. 2.1

SEQ ID NO 7

LENGTH: 1504

TYPE: PRT

ORGANISM: Homo sapiens

FEATURE:

OTHER INFORMATION: DATABASE ACCESSION NUMBER: Genbank,

OTHER INFORMATION: accession number: NP_009115

US-09-932-145-7

Query Match

Best Local Similarity 5.0%; Score 125; DB 9; Length 1504;

Matches 102; Conservative 65; Mismatches 165; Indels 148; Gaps 25;

QY 23 GIKPGIV-HFGVGNFFRAH--EAFYVEQILEHAPDWAIVGVGLTGSRSKKAEEFKAQD 79

Db 105 GTPRVLAHLHFHYEINGITAAAELEFKEBQ--LLGAG-----EVAIGP 151
Qy 80 C-LYSLTETAPSGKSTVRVMGALRDYLLAPADPAVLKHLVDPALRIVSMTI--TEGGY- 135
Db 152 LQLTAVTEQLOOQAPT-----CASGDATDGLHILDTFCLRYLKYKVSGETGPGF 200
Qy 136 --NINETTGAFDLENAAVKADLNKPEKPSVFGYVVEALRRRDAGGKAFVMSCDNLRH 193
Db 201 TSNIQEQLLPFDLS-----IFKSLHQ-----VEISHCDAKHI 232
Qy 194 NGVARKAFGLYAKAR-----DPELAKWIEENATFPNGMVDRTITPTVSAEIA 240
Db 233 RGLVASPTLATLSVRSFATSMSKEVLVPEASEFDEWEPEGTTL-EGPVTAVIPTWQALTT 291
Qy 241 KKL--NAASGLDDDLPLVAEDFHWVLEDOFADGRPPLEKAGVOMGVG--TDW 290
Db 292 LDLSHNSISESVKLIP-----KIEFLD-----LSHNGLLVDNLQHLVNLVHLDL 339
Qy 291 EYVK-----IRMLN-AGHYMLCFFGI-----LVGYENVDDAIEDSELLGNLK 331
Db 340 SYNKLSSLEGLHTKLNKIKTLNLAGNLESLSGLHKLYSLVNLDRNRLEQMEEVRSIG 399
Qy 332 NYLKNVDIPTLKAPSGMTLEGYRDSVISRFSNKAMS---DOTLRIADGCGSKVQVFWTET 388
Db 400 SLPCLHVSLLNNPLSI-IPDYRTKVLQAQGERASEVCLDDTVTTEKE-LDTVEVL--KA 455
Qy 389 VRRRAIEDKRLDSRIAFAGSIAYLEMLRGRDEKG-----TYESSEPTYGDAEWKL 437
Db 456 IQAKAEVSKLS-----NPEKKGEDSRLSAAPCIRPSSPPTVAPASASL 501

RESULT 8
US-09-824-574-3
; Sequence 3, Application US/09824574
; Publication No. US20030077800A1
; GENERAL INFORMATION:
; APPLICANT: Rouleau, Natalie
; APPLICANT: Mollanen, Anu-Maarit
; APPLICANT: Palvimio, Jorma J.
; APPLICANT: Jone, Olli A.
; TITLE OF INVENTION: AR14 Gene and Protein
; FILE REFERENCE: 2630-109
; CURRENT APPLICATION NUMBER: US/09/824,574
; CURRENT FILING DATE: 2001-04-03
; NUMBER OF SEQ ID NOS: 7
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 3
; LENGTH: 1703
; TYPE: PRT
; ORGANISM: Saccharomyces cerevisiae
US-09-824-574-3

Query Match 4.3%; Score 107; DB 9; Length 1703;
Best Local Similarity 21.5%; Pred. No. 2.5;
Matches 98; Conservative 75; Mismatches 168; Indels 114; Gaps 24;

Qy 45 VEQILEHAPDWAIVGVGLTSGDSRKKAEFEKADCLYSLTETAPSG-----KSTVRV 97
Db 280 IQKSIHPPDFKRLMLLSLSEFARRRQPTDQ--NQSNLNGGNTTOOQGTNSHNTNTDNV 338
Qy 98 MGALRDYLL-----APADPEAV---LKHLYDPAIRIVS---MTITEGGYNINETT 141
Db 339 SGLTRNAPLDSKDENFASVSPAGSSVHNAKNGTLDKNSQTSVGTPTQTESKEENET- 397
Qy 142 GAFDLENAVKA-----DLKNPEKPSVFGYVVEALRRRDAGGKAFVMSCD----- 189
Db 398 ----ISNVAKTAPNSKNTHTTEQNNPPKPKQ--PVLNVLDQDYKEGIVKVDIDDPDMVD 451
Qy 190 ----NLRHNGNVARKAFLGVAKARDPELAKWIEENATFPNGMVDRTITPTVSAEIAKLN 244
Db 452 SFTMPNISHS-NIDYQTLA-----NSDHAKFTIEPGVLPVG-IDTHTATDIYOTLIALN 504

Qy 245 AASGLDDDLPLV-----AEDFHOWVLEDOFADGRPPLEKAGVOMGVGDTWWEYVKIRMLN 299
Db 505 LDTTVNCLOKLLNDECTSTRENALYDYALQLPLQKA---VRGHVLOFENHONSLLT 561
Qy 300 AGHYMLCFFPGLVGYENVDDAIEDSELLGNLKNYLNKDVIPITLKAPSGMTLEGYRDSVIS 359
Db 562 NTH-----PNFLSKIRIN--VODALLTNQL--YKNHEL-----LKLERKKTEAVA 603
Qy 360 RFS--NKAMSDOTLRIASDGCSKVQVFWTETVRRRAIEDKRLDSRIAEG---IASYLEMLR 414
Db 604 RLKSMNKSALNQYNR-----RODKN-KRLKFGHRLIATHHTLER 642
Qy 415 GRDEKGYTESSEPTYGDAEWKLAKADDFESSKL 449
Db 643 D-EQKRAEKRAKE-----RLQALKANDEEAYIKL 670

RESULT 9

US-09-801-368-340
; Sequence 340, Application US/09801368
; Patent No. US20020128250A1
; GENERAL INFORMATION:
; APPLICANT: Busby, Robert
; APPLICANT: Cali, Brian
; APPLICANT: Hecht, Peter
; APPLICANT: Holtzman, Doug
; APPLICANT: Madgen, Kevin
; APPLICANT: Maxon, Mary
; APPLICANT: Milne, Todd
; APPLICANT: NO. US20020128250A1man, Thea
; APPLICANT: Royer, John
; APPLICANT: Salama, Sofie
; APPLICANT: Sherman, Amir
; APPLICANT: Silvers, Jeff
; APPLICANT: Summers, Eric
; TITLE OF INVENTION: Methods for Improving Secondary Metabolite Production in Fungi
; FILE REFERENCE: 109272.147
; CURRENT APPLICATION NUMBER: US/09/801,368
; CURRENT FILING DATE: 2001-03-07
; PRIOR APPLICATION NUMBER: US 09/487,558
; PRIOR FILING DATE: 2000-01-19
; PRIOR APPLICATION NUMBER: US 60/160,587
; PRIOR FILING DATE: 1999-10-20
; NUMBER OF SEQ ID NOS: 440
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 340
; LENGTH: 1703
; TYPE: PRT
; ORGANISM: Saccharomyces cerevisiae
US-09-801-368-340

Query Match 4.3%; Score 107; DB 10; Length 1703;
Best Local Similarity 21.5%; Pred. No. 2.5;
Matches 98; Conservative 75; Mismatches 168; Indels 114; Gaps 24;

Qy 45 VEQILEHAPDWAIVGVGLTSGDSRKKAEFEKADCLYSLTETAPSG-----KSTVRV 97
Db 280 IQKSIHPPDFKRLMLLSLSEFARRRQPTDQ--NQSNLNGGNTTOOQGTNSHNTNTDNV 338
Qy 98 MGALRDYLL-----APADPEAV---LKHLYDPAIRIVS---MTITEGGYNINETT 141
Db 339 SGLTRNAPLDSKDENFASVSPAGSSVHNAKNGTLDKNSQTSVGTPTQTESKEENET- 397
Qy 142 GAFDLENAVKA-----DLKNPEKPSVFGYVVEALRRRDAGGKAFVMSCD----- 189
Db 398 ----ISNVAKTAPNSKNTHTTEQNNPPKPKQ--PVLNVLDQDYKEGIVKVDIDDPDMVD 451
Qy 190 ----NLRHNGNVARKAFLGVAKARDPELAKWIEENATFPNGMVDRTITPTVSAEIAKLN 244
Db 452 SFTMPNISHS-NIDYQTLA-----NSDHAKFTIEPGVLPVG-IDTHTATDIYOTLIALN 504
Qy 245 AASGLDDDLPLV-----AEDFHOWVLEDOFADGRPPLEKAGVOMGVGDTWWEYVKIRMLN 299

Db 505 LDTTVNDCCLKLLNDECTESTRENALDYALQLLPQKA---VRGHVLQFEWHQNSLLT 561
Qy 300 AGHVMLCPGILVGVENVDATEDSELGNLKNYLNKDVIPTLKAPSGMTLEGYRDSVIS 359
Db 562 NTH-----PNFLSKTRNIN--VODALLNQL--YKNHEL-----LKLKKTEAVA 603
Qy 360 RFS--NKAMSDQTLRIASDGCCKVQVFWTETVRRRAIEDKRLSRIAIFG---IASYLEMLR 414
Db 604 RLKSMKSAINOYR-----RQDKKN-KRLKFGHRLIATHTNLER 642
Qy 415 GRDEKGGYSESEPTYGDAEWKLAKADDFESSLKL 449
Db 643 D-EQRAEKAKE-----RLQALKANDEEAYIKL 670

RESULT 10

US-09-860-846-35

; Sequence 35, Application US/09860846
; Patent No. US20020164742A1
; GENERAL INFORMATION:
; APPLICANT: Sherman, D.H.
; APPLICANT: Liu, H.
; APPLICANT: Xue, Y.
; APPLICANT: Zhao, L.
; TITLE OF INVENTION: DNA encoding methymycin and pikromycin
; FILE REFERENCE: 600.438US1
; CURRENT APPLICATION NUMBER: US/09/860.846
; CURRENT FILING DATE: 2001-05-18
; PRIOR APPLICATION NUMBER: 09/105,537
; PRIOR FILING DATE: 1998-06-26
; SOFTWARE: FastSeq for Windows Version 3.0
; NUMBER OF SEQ ID NOS: 43
; SEQ ID NO 35
; LENGTH: 1562
; TYPE: PRT
; ORGANISM: Streptomyces venezuelae
US-09-860-846-35

Query Match 4.2%; Score 105.5; DB 9; Length 1562;
Best Local Similarity 20.8%; Pred. No. 2.9;
Matches 100; Conservative 74; Mismatches 194; Indels 113; Gaps 21;
Qy 87 TAPGKSTVRVM-GALRDYLLAPADPEAVLKH-----LVDPAIRIVSMTITEGGYNINET 140
Db 313 TAPHGPSQORVIRRALADARLAPGDVVEAHGTCTRLGDP-IEAQAIIATYVG-----QEK 367
Qy 141 TGAFDLENAVNAKADLNKEKSTVFGYV--VEALRR-----RWDAGCKA 182
Db 368 SSEQPLRLGALKSNIGHTQAAAGVAGVVKMVOAMRHGLLPKTLHVDEPSDQIDNSAGTVE 427
Qy 183 FTVMSCD-NLRHNGNVARKAFGLYAKARDP-----ELAKWTEENATF--PNG-----MV 228
Db 428 LITEAVDMPKODGGLRRAAVSSFGISGTNAHVLEEAPEAVEDSPAVEPPAGGVVPPV 487
Qy 229 DRITP-TVSAEIAKLNAAAGLDDDLPLVAEDFHOWLEDQFADGRPPLEKAGVQMVGDV 287
Db 488 SAKTPAALDAQIGQLAAYADGRTDVPDPAV-----ARALVDSRTAMEHRAV-AVGDS 538
Qy 288 TDWEYVKTRM-----LNAGHVMLCFPG-----ILVGVENVDDATEDSELGNLKNY 333
Db 428 LITEAVDMPKODGGLRRAAVSSFGISGTNAHVLEEAPEAVEDSPAVEPPAGGVVPPV 487
Qy 229 DRITP-TVSAEIAKLNAAAGLDDDLPLVAEDFHOWLEDQFADGRPPLEKAGVQMVGDV 287
Db 488 SAKTPAALDAQIGQLAAYADGRTDVPDPAV-----ARALVDSRTAMEHRAV-AVGDS 538
Qy 288 TDWEYVKTRM-----LNAGHVMLCFPG-----ILVGVENVDDATEDSELGNLKNY 333
Db 539 REALDALRMEPEGLVRGTSSDVGRVAFVFPQGTQWAGMGAEILLDSSPEFAASMAECETA 598
Qy 334 LNKDVIPITKA-----PSGMTLE-----GYRDSVISRFSNKAMSDQTL----- 371
Db 599 LSRVYDWSLEAVVREPQCAPTLDRVDVVPVTFVAVVMSLAKVWQHGGITPQAVVCHSOG 658
Qy 372 -----RIASDGCCKVQVFWTETVRRRAIEDKRLSRIAIFGSIASYLEMLRGRDEK--- 420
Db 659 IAAAYVAGALTLDAAARVVTLRKSKSIAAHLAGKGMISLALDEAAVLRSLDFDGLSVAA 718
Qy 421 -----GTVESSEPTYGDAEWKLAKADDFESSLKLPAFDGWRDLDTSELDOKVIVLRKIIR 475
Db 719 VNGPTATVVGSDPTQIEELARTCEADGVRARI-IP-----VDYASHRSQVEIIEKELA 770

Qy 476 E 476

Db 771 E 771

RESULT 11

US-09-988-384B-35
; Sequence 35, Application US/0988384B
; Publication No. US20030073824A1
; GENERAL INFORMATION:
; APPLICANT: Sherman, D.H.
; APPLICANT: Liu, H.
; APPLICANT: Xue, Y.
; APPLICANT: Zhao, L.
; TITLE OF INVENTION: DNA encoding methymycin and pikromycin
; FILE REFERENCE: 600.536US1
; CURRENT APPLICATION NUMBER: US/09/988.384B
; CURRENT FILING DATE: 2001-11-19
; PRIOR APPLICATION NUMBER: PCT/US99/14398
; PRIOR FILING DATE: 1999-06-25
; PRIOR APPLICATION NUMBER: US 09/105,537
; PRIOR FILING DATE: 1998-06-26
; NUMBER OF SEQ ID NOS: 53
; SEQ ID NO 35
; LENGTH: 1562
; TYPE: PRT
; ORGANISM: Streptomyces venezuelae
US-09-988-384B-35

Query Match 4.2%; Score 105.5; DB 9; Length 1562;
Best Local Similarity 20.8%; Pred. No. 2.9;
Matches 100; Conservative 74; Mismatches 194; Indels 113; Gaps 21;
Qy 87 TAPGKSTVRVM-GALRDYLLAPADPEAVLKH-----LVDPAIRIVSMTITEGGYNINET 140
Db 313 TAPHGPSQORVIRRALADARLAPGDVVEAHGTCTRLGDP-IEAQAIIATYVG-----QEK 367
Qy 141 TGAFDLENAVNAKADLNKEKSTVFGYV--VEALRR-----RWDAGCKA 182
Db 368 SSEQPLRLGALKSNIGHTQAAAGVAGVVKMVOAMRHGLLPKTLHVDEPSDQIDNSAGTVE 427
Qy 183 FTVMSCD-NLRHNGNVARKAFGLYAKARDP-----ELAKWTEENATF--PNG-----MV 228
Db 428 LITEAVDMPKODGGLRRAAVSSFGISGTNAHVLEEAPEAVEDSPAVEPPAGGVVPPV 487
Qy 229 DRITP-TVSAEIAKLNAAAGLDDDLPLVAEDFHOWLEDQFADGRPPLEKAGVQMVGDV 287
Db 488 SAKTPAALDAQIGQLAAYADGRTDVPDPAV-----ARALVDSRTAMEHRAV-AVGDS 538
Qy 288 TDWEYVKTRM-----LNAGHVMLCFPG-----ILVGVENVDDATEDSELGNLKNY 333
Db 539 REALDALRMEPEGLVRGTSSDVGRVAFVFPQGTQWAGMGAEILLDSSPEFAASMAECETA 598
Qy 334 LNKDVIPITKA-----PSGMTLE-----GYRDSVISRFSNKAMSDQTL----- 371
Db 599 LSRVYDWSLEAVVREPQCAPTLDRVDVVPVTFVAVVMSLAKVWQHGGITPQAVVCHSOG 658
Qy 372 -----RIASDGCCKVQVFWTETVRRRAIEDKRLSRIAIFGSIASYLEMLRGRDEK--- 420
Db 659 IAAAYVAGALTLDAAARVVTLRKSKSIAAHLAGKGMISLALDEAAVLRSLDFDGLSVAA 718
Qy 421 -----GTVESSEPTYGDAEWKLAKADDFESSLKLPAFDGWRDLDTSELDOKVIVLRKIIR 475
Db 719 VNGPTATVVGSDPTQIEELARTCEADGVRARI-IP-----VDYASHRSQVEIIEKELA 770

Qy 476 E 476

Db 771 E 771

RESULT 12

US-09-861-289-35

```

; CURRENT APPLICATION NUMBER: US/09/860,846
; CURRENT FILING DATE: 2001-05-18
; PRIOR APPLICATION NUMBER: 09/105,537
; PRIOR FILING DATE: 1998-06-26
; NUMBER OF SEQ ID NOS: 43
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 6
; LENGTH: 11877
; TYPE: PRT
; ORGANISM: Streptomyces venezuelae
; US-09-860-846-6

Query Match
Best Local Similarity 4.2%; Score 105.5; DB 9; Length 11877;
Matches 100; Conservative 74; Mismatches 194; Indels 113; Gaps

QY 87 TAPSGKSTVRVM-GALRDYLLAPADPEAVLKH-----LVDPAIRIVSVTITEGGYNINET 140
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 9001 TAPHGSPQQRVTRRALADARLARLPAGVDVVEAHGTCGTRLGDP-IEAQAALIATYG---QEK 9055
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 141 TGAFDLENAAVKADLNKPKSTVFYGV--VEALRR-----RWDAGGKA 182
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 9056 SSQOPRLGALKSNIGHTQAAAGVAGVKNWQAMRHGLLPKTLHVDEPSQIDWSAGTVE 9115
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 183 FTYMSCD-NLRHNGNVYARKAFLGYAKARDP-----ELAKWIEENATF--PNG-----MV 228
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 9116 LLTEAVDWPEKQDGLRRAAVSSFGISGTNAHVHLEEAPEAVEDSPAEPAGGGVVPWPV 9175
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 229 DRTPP-TVSAETAKLNAASGLDDDLPLVAEDFHOWILEDQFAGRPPLPKAGVQMVDV 287
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 9176 SAKTPAALDAQIGQLAAYADGRDTDVA-----ARALVDSRTAMEHRAV-AVGDS 9226
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 288 TDWEYVKIRM-----LNAGHWMLCPG-----ILVGYENVDDATEDSELGLNKNY 333
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 9227 REALRDALRMPEGLVRGTSDDVRGFAVFPFGGTQWAGMGAEILLDSSPEFAASMAECETA 9286
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 334 LNKRDVPTLKA-----PSGMTLE-----GYRDSVISRFSNKAMSQDTL----- 371
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 9287 LSRVDMSEAVVYRQEPGAPTLDRVDVQVPTFAVMVSLAKVWQHGGITPQAVVGHSGQE 9346
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 372 -----RIADGCKSVQVFTETVRRADIEDKRLSRIAFGSIASYLEMLRGREKG--- 420
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 9347 IAAAYVAGALTLDLDAARVVYTLRSKSTAAHLGCKGMI SLALDEAAVLKRLSDFDGLSVAA 9406
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 421 ----CTYESSEPTYGDAEWKLAKADDFESSLKLPAFDGWRDLDTSELDDQKVVIVLRKIIR 475
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 9407 VNGPTATVVGSGPTQTEELARTCEADGVRARI-IP-----VDYASHSRQVEIIEIKELA 9458
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 476 E 476
|
Db 9459 E 9459

RESULT 14
US-09-861-289-6
; Sequence 6, Application US/09861289
; Patent No. US20020110897A1
; GENERAL INFORMATION:
; APPLICANT: Sherman, D.H.
; APPLICANT: Liu, H.
; APPLICANT: Xue, Y.
; APPLICANT: Zhao, L.
; TITLE OF INVENTION: DNA encoding methymycin and pikromycin
; FILE REFERENCE: 600.438US1
; CURRENT APPLICATION NUMBER: US/09/861,289
; CURRENT FILING DATE: 2001-05-18
; PRIOR APPLICATION NUMBER: 09/105,537
; PRIOR FILING DATE: 1998-06-26
; NUMBER OF SEQ ID NOS: 43
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 6
; LENGTH: 11877
; TYPE: PRT

```

ORGANISM: Streptomyces venezuelae
US-09-861-289-6

Query Match 4.2%; Score 105.5; DB 10; Length 11877;
Best Local Similarity 20.8%; Pred. No. 57;
Matches 100; Conservative 74; Mismatches 194; Indels 113; Gaps 21;

QY 87 TAPSGKSTVRVM-GALRDYLLAPADPEAVLKH-----LVDPAIRIVSMITTEGGYNINET 140
DB 9001 TAPHGPSQORVIRRALADARLAPGDVVDVEAHGTTGLGDP-IEAQALIATYG-----OEK 9055

QY 141 TGAFDLENAAKADLNKPEKSTVFYGV--VEALRR-----RWDAGGKA 182
DB 9056 SSEQPLRLGALKSNIGHTQAAAAGVAGVIKMQAMRHGLLPTKTLHVDEPSDQIDWSAGTVE 9115

QY 183 FTVMSCD-NLRHNGNVARKAFGLYAKARDP-----ELAKWIEENATF--PNG-----MV 228
DB 9116 LITEAVDMPEKODGGLRRAAVSSFGISGTNAHVLEEAPEAVEDSPAEPAGGGVVPWPV 9175

QY 229 DRITP-TVSAETIAKKLNAAAGLDDDLPLVAEDFHQWLEDFADGRPPLEKAGVOMGDV 287
DB 9176 SAKTAAALDAIQGLAAVADGRTDVPDPAV-----ARALVDSRTAMEHRAV-AVGDS 9226

QY 288 TDWEYVVKIRM-----LNAGHVMLCFPG-----ILVGYENVDDAIEDSELLGNLKNY 333
DB 9227 REALRDALRMPEGLVRGTSSDVGRAVAFVFPQGTQWAGMGAEELDSPEFAASMAECETA 9286

QY 334 LNKDVIPITLKA-----PSGMTLE-----GYRDSVISRFSNKAMSDOTL----- 371
DB 9287 LSRVYDWSLEAVVROEFGAPTLDRVDVYQVPTFAVMVSLAKVWQHGGITPOAVVCHSOG 9346

QY 372 -----RIASDGCCKVQVFWETVRRRAIEDKRLSRIAFIGIASYLEMLRGRDEKG--- 420
DB 9347 IAAAYVAGALTLDAAARVVTLSKSIAAHLAKGKGMISLALDEAAVLKRLSDFDGLSVAA 9406

QY 421 -----GTVESSEPTYGDAEWKLAADDFESSILKLPAPDGRDLDTSELDDQKVIVLRLKIIR 475
DB 9407 VNGPTATVVGSDPTQIEELARTCEADGVRARI-IP-----VDYASHRSQVEIIEKELA 9458

QY 476 E 476
DB 9459 E 9459

RESULT 15
US-09-988-384B-6
; Sequence 6, Application US/09988384B
; Publication No. US20030073824A1
; GENERAL INFORMATION:
; APPLICANT: Sherman, D.H.
; APPLICANT: Liu, H.
; APPLICANT: Xue, Y.
; APPLICANT: Zhao, L.
; TITLE OF INVENTION: DNA encoding methymycin and pikromycin
; FILE REFERENCE: 600.536US1
; CURRENT APPLICATION NUMBER: US/09/988.384B
; CURRENT FILING DATE: 2001-11-19
; PRIOR APPLICATION NUMBER: PCT/US99/14398
; PRIOR FILING DATE: 1999-06-25
; PRIOR APPLICATION NUMBER: US 09/105,537
; PRIOR FILING DATE: 1998-06-26
; NUMBER OF SEQ ID NOS: 53
; SEQ ID NO 6
; LENGTH: 12199
; TYPE: PRT
; ORGANISM: Streptomyces venezuelae
US-09-988-384B-6

Query Match 4.2%; Score 105.5; DB 9; Length 12199;
Best Local Similarity 20.8%; Pred. No. 59;
Matches 100; Conservative 74; Mismatches 194; Indels 113; Gaps 21;

QY 87 TAPSGKSTVRVM-GALRDYLLAPADPEAVLKH-----LVDPAIRIVSMITTEGGYNINET 140

DB 9323 TAPHGPSQORVIRRALADARLAPGDVVDVEAHGTTGLGDP-IEAQALIATYG-----OEK 9377

QY 141 TGAFDLENAAKADLNKPEKSTVFYGV--VEALRR-----RWDAGGKA 182
DB 9378 SSEQPLRLGALKSNIGHTQAAAAGVAGVIKMQAMRHGLLPTKTLHVDEPSDQIDWSAGTVE 9437

QY 183 FTVMSCD-NLRHNGNVARKAFGLYAKARDP-----ELAKWIEENATF--PNG-----MV 228
DB 9438 LITEAVDMPEKODGGLRRAAVSSFGISGTNAHVLEEAPEAVEDSPAEPAGGGVVPWPV 9497

QY 229 DRITP-TVSAETIAKKLNAAAGLDDDLPLVAEDFHQWLEDFADGRPPLEKAGVOMGDV 287
DB 9498 SAKTAAALDAIQGLAAVADGRTDVPDPAV-----ARALVDSRTAMEHRAV-AVGDS 9548

QY 288 TDWEYVVKIRM-----LNAGHVMLCFPG-----ILVGYENVDDAIEDSELLGNLKNY 333
DB 9549 REALRDALRMPEGLVRGTSSDVGRAVAFVFPQGTQWAGMGAEELDSPEFAASMAECETA 9608

QY 334 LNKDVIPITLKA-----PSGMTLE-----GYRDSVISRFSNKAMSDOTL----- 371
DB 9609 LSRVYDWSLEAVVROEFGAPTLDRVDVYQVPTFAVMVSLAKVWQHGGITPOAVVCHSOG 9668

QY 372 -----RIASDGCCKVQVFWETVRRRAIEDKRLSRIAFIGIASYLEMLRGRDEKG--- 420
DB 9669 IAAAYVAGALTLDAAARVVTLSKSIAAHLAKGKGMISLALDEAAVLKRLSDFDGLSVAA 9728

QY 421 -----GTVESSEPTYGDAEWKLAADDFESSILKLPAPDGRDLDTSELDDQKVIVLRLKIIR 475
DB 9729 VNGPTATVVGSDPTQIEELARTCEADGVRARI-IP-----VDYASHRSQVEIIEKELA 9780

QY 476 E 476
DB 9781 E 9781

Search completed: May 8, 2003, 16:31:27
Job time : 91 secs

GenCore version 5.1.4_p5_4578
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: May 8, 2003, 16:29:57 ; Search time 46 Seconds
(without alignments)
1013.591 Million cell updates/sec

Title: US-09-926-163B-2
Perfect score: 2516
Sequence: 1 MITRETLKSLPANVQAPPDY.....KVIVLRKIIREKGVKAAIPA 485
Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 283224 seqs, 96134422 residues
Total number of hits satisfying chosen parameters: 283224

Minimum DB seq length: 0
Maximum DB seq length: 2000000000
Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : PIR_73:*
1: pir1:*
2: pir2:*
3: pir3:*
4: pir4:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	922	36.6	491	2	D83353
2	789	31.4	502	2	S50519
3	779	31.0	487	2	A96022
4	775.5	30.8	486	2	B91289
5	775.5	30.8	486	2	E86130
6	774.5	30.8	486	2	S56548
7	771.5	30.7	486	2	A64909
8	769.5	30.6	486	2	G90897
9	768	30.5	478	2	T03548
10	763.5	30.3	486	2	H85719
11	759.5	30.2	488	2	A00679
12	753.5	29.9	490	2	A01156
13	746.5	29.7	488	2	F91011
14	746.5	29.7	488	2	B95856
15	739.5	29.4	488	2	C64986
16	730	29.0	490	2	A00893
17	730	29.0	494	2	A63103
18	730	29.0	525	2	C98183
19	721	28.7	458	2	F87433
20	700.5	27.8	502	2	A53569
21	650.5	25.9	490	2	A59293
22	650.5	25.9	490	2	A02990
23	569	22.6	463	2	A0283
24	371	14.7	539	2	H72422
25	365.5	14.5	544	2	H86828
26	343.5	13.7	482	2	E96985
27	331	13.2	480	2	C59853
28	299.5	11.9	512	2	D83711
29	288	11.4	483	2	A80072

ALIGNMENTS

RESULT 1

D83353
mannitol dehydrogenase PA2342 [imported] - Pseudomonas aeruginosa (strain PA01)
C:Species: Pseudomonas aeruginosa
C:Date: 15-Sep-2000 #sequence_revision 15-Sep-2000 #text_change 31-Dec-2000
C:Accession: D83353
R:Stover, C.K.; Pham, X.Q.; Erwin, A.L.; Mizoguchi, S.D.; Warren, P.; Hickey, M.J.;
adman, S.; Yuan, Y.; Brody, L.L.; Coulter, S.N.; Folger, K.R.; Kas, A.; Larbig, K.; L
.: Lory, S.; Olson, M.V.
Nature 406, 959-964, 2000
A:Title: Complete genome sequence of Pseudomonas aeruginosa PA01, an opportunistic pa
A:Reference number: A82950; MUID:20437337; PMID:10984043
A:Accession: D83353
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-491 <STO>
A:Cross-references: GB:AE004660; GB:AE004091; NID:99948372; PIDN:AAG05730.1; GSPDB:GN
A:Experimental source: strain PA01
C:Genetics:
A:Gene: mtID; PA2342
C:Superfamily: conserved hypothetical protein YEL070W

Query Match 36.6%; Score 922; DB 2; Length 491;
Best Local Similarity 45.2%; Pred. No. 3.3e-57;
Matches 199; Conservative 66; Mismatches 171; Indels 4; Gaps 2;
Qy 2 ITRETLKSLPANVQAPPDYIDIGIKPGIVHFGVGNFFRAHEAFYVEQILE--HAPDMAIVG 59
Db 3 LNROHLPLLATAVARPSYDPAQLRQGIHVIGVGFHRAHQAAYTDALMNRGALDWAICG 62
Qy 60 VGLTGSRSRKKAEFEKADCLYSLTETAPSGKSTVRVGMALRDYLLAPADPEAVLKHLV 119
Db 63 AGLRSDDRAMHDA--LAAQDYLYTLIELGDQDPDTEVRVIGALSGLMLAEDGAELKLA 120
Qy 120 DPAIRIVSMITTEGGYNINETTGAFDLENAAYKADLKNKEPSTVFYGVVEALRRWDAG 179
Db 121 EPAIRIVSLITTEGGYCIDGSGEFLAEPLVLRHDLANPRTPRGVFGFLCEALRRRTG 180
Qy 180 GKAFVTMSCDNLRHNGNVARKAFGLYAKARDELAKEATEENATPPNGMVDRITPTVSAEI 239
Db 181 VPAFTVMSCDNLPHNGEVARKALLAFAERLDPLGLARWIATHVSFPFNAMVDRTPTSPAH 240
Qy 240 AKKLNAAAGLDDDLPLVAEDFHQWLEDFADGRPPLEKAGVQWGVGDYTDMEYKIRMLN 299
Db 241 RRLQAQRHEDVEDAMPVCEPFVQVWLEDFRSAGRPAWEKVGQVQFTDDVTPEEMKIGLLN 300
Qy 300 AGHVLMCEPGLVGVYENVDDAIEDSELNKNYLNKQVITPLKAPSGMTLEGYRDSVIS 359
Db 301 GSHLALTYLGLRGYRFVHETLGDPLLRVYVAFHMDRVDAPLLAPVPGIDILERYKDSLVE 360
Qy 360 RFSNKAUSDQTLRIASDCGSKVQVFVETVTRRAIEDKRLSRIAFGLSYLEMLRGRDEK 419

altronate oxidored
tagatone oxidored
tagatone reduc
probable sigma fac
mannitol-1-phospha
mannitol-1-phospha
mannitol-1-phospha
mannitol-1-phospha
hypothetical prote
mannitol-1-phospha
mannitol-1-phospha
mannitol-1-phospha
mannitol-1-phospha
mannitol-1-phospha

RESULT 3

A96022
probable fructuronate reductase (EC 1.1.1.57) [imported] - Sinorhizobium meliloti
C:Species: Sinorhizobium meliloti
C:Date: 24-Aug-2001 #sequence_revision 24-Aug-2001 #text_change 14-Sep-2001
C:Accession: A96022
R:Finan, T.M.; Weidner, S.; Wong, K.; Buhrmester, J.; Chain, P.; Vorholter, P.; Probst, A.; Weidner, S.; U.S.A. 98, 9889-9894, 2001
A:Title: The complete sequence of the 1.683-kb pSymB megaplasmid from the
A:Reference number: A95842; MUID:21396508; PMID:11481431
A:Accession: A96022
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-487 <KUR>
A:Cross-references: GB:AL591985; PIDN:CAC49841.1; PID:g15141329; GSPDB:GN5B
A:Experimental source: Strain 1021, megaplasmid pSymB
R:Galibert, F.; Finan, T.M.; Long, S.R.; Puller, A.; Abolla, P.; Ampe, F.; Pella, D.; Chain, P.; Cowie, A.; Davis, R.W.; Dreano, S.; Federspiel, N.A.; L.; Hyman, R.W.; Jones, T.
Science 293, 668-672, 2001
A:Authors: Kahn, D.; Kahn, M.L.; Kalman, S.; Keating, D.H.; Kiss, E.; Kombeault, P.; Vandenbol, M.; Vorholter, F.J.; Weidner, S.; Wells, D.H.; Wolcott, A.; Title: The composite genome of the legume symbiont Sinorhizobium meliloti
A:Reference number: A96039; MUID:21368234; PMID:11474104
A:Contents: annotation
C:Genetics:
A:Gene: uxuB; Smb20749
A:Genome: plasmid
C:Superfamily: conserved hypothetical protein YEL070w
C:Keywords: oxidoreductase

[illegible]

Db 3 LSRKTIIDRLPRTVKRPHYDLGTVTGIVHLGIGAFHRAHQAVYTDGLLSEDPGICGV 62
Qy 62 LTGDSRSKKAEEFKAQDCLYSLTETAPSGKSTVRVMGALROYLLAPADPEAVLKLHLYDP 121
Db 63 L---RSPETRDALHPDGLYTLAVQDGE-SELSVGSVVELLCAPDDPEAVLRRMADP 117
Qy 122 AIRIVSMITTEGGYNINETTGAFLDLENAAVKADLNKPKPSTVFGVYVEALRRRWADGCK 181
Db 118 GTRIVSLRTEKGYCHNPATGTLDEGHPDIVHDLANPARPSRGAIGFIVEAISRRYSAGIA 177
Qy 182 AFTVMSCDNLRHNGVARKAFGLGAKARDELAKWIEENATFPNGMVDRIPTVSAETAK 241
Db 178 PFTLSCDNLPGHVLKARIVTQFAEARDPALAA-VVRNVPSTMDVRIPTATDSRS 236
Qy 242 KLNASGLDDDLPLVAEDFHQWLEDDQADGRPPLEKAGVQMGVDVTWYVVKIRMLNAG 301
Db 237 AVASAMGLEDAWPIINTEPFROWIIEEDPLGRPAWEKAGALFVQDSAFEFMKLRLLNGS 296
Qy 302 HVMLCFPGILVGYENVDDAIEDSELLGNLKNVLDKVIPTTLKAPSGMTLEGYRDSVISRFP 361
Db 297 HSTLAYLGLAGAEVADAMALAGMEALVEGLMRHEVSTPLPELPGFDLPAYRAELQRF 356
Qy 362 SNKMSDQTLRIASDGCCKVQVFWETVTRRAIEDKRLSRIAFAFGIASYLEMLRGRDEKGG 421
Db 357 RNPALRHRTWQIAMDGSOKLPORLLGSTRDLQAGAGYDRLLALGVAWWMRYARGLDEAGR 416
Qy 422 TYESEP 428
Db 417 PIDVRDP 423
RESULT 4
B-mannanase oxidoreductase EC5282 [imported] - Escherichia coli (strain O157:H7, substr
C:Species: Escherichia coli
C:Date: 18-Jul-2001 #sequence_revision 18-Jul-2001 #text_change 03-Aug-2001
C:Accession: B91289
R:Hayashi, T.; Makino, K.; Ohnishi, M.; Kurokawa, K.; Ishii, K.; Yokoyama, K.; Han, C.G.
gasawara, N.; Yasunaga, T.; Kuhara, S.; Shiba, T.; Hattori, M.; Shinagawa, H.
DNA Res. 8, 11-22, 2001
A:Title: Complete genome sequence of enterohemorrhagic Escherichia coli O157:H7 and gene
A:Reference number: A99629; MUID:21156231; PMID:11258796
A:Accession: B91289
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-486 <HAY>
A:Cross-references: GB:BA000007; PIDN:BA038705.1; PID:q13364760; GSPDB:GN00154
A:Experimental source: strain O157:H7, substrain RMD 0509952
C:Genetics:
A:Gene: EC5282
C:Superfamily: conserved hypothetical protein YEL070w
Query Match 30.8%; Score 775.5; DB 2; Length 486;
Best Local Similarity 36.4%; Pred. No. 6.8e-47;
Matches 177; Conservative 87; Mismatches 205; Indels 17; Gaps 9;
Qy 9 SLPANVOAPPYDIDGKPGIVHFGVGNFRAHEAFYVEQILEHA-PDWAIVGVGL-TGSD 66
Db 8 NLP--VARPSWDHRSRLESRIVHLGCGAFHRAHQALYTHLLESTDSMGICEVNLMPGND 65
Qy 67 RSKKKAEEFKAQDCLYSLTETAPSGKSTVRVMGALROYLLAPADP-EAVLKLHLYDPATRI 125
Db 66 R-VLIENLKKQQLLYTVAERG-AESTELKIIGSMKEALHPEIDGCEGILNAMARPTAI 122
Qy 126 VSMITTEGGYNINETTGAFLDLENAAVKADLNKPKPSTVFGVYVEALRRRWADGAKFTV 185
Db 123 VSLTVTEKGYCADAAAGOLDLNNPLIKHDLNPTAPSAIGYIVEALRLRREKGLKFTV 182
Qy 186 MSCDNLRHNGVARKAFGLGAKARDELAKWIEENATFPNGMVDRIPTVSAETAKLNA 245
Db 183 MSCDNLRHNGVARKAFGLGAKARDELAKWIEENATFPNGMVDRIPTVSAETAKLNA 242
Qy 246 ASGLDDDLPLVAEDFHQWLEDDQADGRPPLEKAGVQMGVDVTWYVVKIRMLNAGHYML 305
Db 243 QLVGVDPCAIAACEFROWIIEEDFVNGRPNWDKVKCAQFVADVVPPEMKLRMLNGSHSFL 302
Qy 306 CFPGLVGYENVDDAIEDSELLGNLKNVLDKVIPTTLKAPSGMTLEGYRDSVISRFSNKA 365
Db 303 AYLGVLGGYETIADTMTNPDYRKAFAFALMMQEQAPTLMSPEGTDLNAYATLLIERFSNPS 362
Qy 366 MSDQTLRIASDGCCKVQVFWETVTRRAIEDKRLSRIAFAFGIASYLEMLRGRDEKGGTYES 425
Db 363 LRHRTWQIAMDGSOKLPORLLDPVRLHLONGGSRHRLALGVAGWMRYTQGVDEQGNADIV 422

Db 243 QLVGVDPCAIAACEFROWIIEEDFVNGRPNWDKVKCAQFVADVVPPEMKLRMLNGSHSFL 302
Qy 306 CFPGLVGYENVDDAIEDSELLGNLKNVLDKVIPTTLKAPSGMTLEGYRDSVISRFSNKA 365
Db 303 AYLGVLGGYETIADTMTNPDYRKAFAFALMMQEQAPTLMSPEGTDLNAYATLLIERFSNPS 362
Qy 366 MSDQTLRIASDGCCKVQVFWETVTRRAIEDKRLSRIAFAFGIASYLEMLRGRDEKGGTYES 425
Db 363 LRHRTWQIAMDGSOKLPORLLDPVRLHLONGGSRHRLALGVAGWMRYTQGVDEQGNADIV 422
Qy 426 SEPTYGDEWLAKADDESSKLPAPFGWDRDLTSELDDQK-----VIVLRKIIREKGV 479
Db 423 VDPML--AEFKINA-QYQAGRVKALLGLSGIFADDPQNADEVGAVTAAYQOOLCERGA 479
Qy 480 KAAIPA 485
Db 480 RECVAA 485
RESULT 5
E86130
mannanase oxidoreductase [imported] - Escherichia coli (strain O157:H7, substrain EDL
C:Species: Escherichia coli
C:Date: 16-Feb-2001 #sequence_revision 16-Feb-2001 #text_change 14-Sep-2001
C:Accession: E86130
R:Perna, N.T.; Plunkett III, G.; Burland, V.; Mau, B.; Glasner, J.D.; Rose, D.J.; May
iller, L.; Grotbeck, E.J.; Davis, N.W.; Lim, A.; Dimalanta, E.; Potamouisis, K.; Apoda
Nature 409, 529-533, 2001
A:Title: Genome sequence of enterohemorrhagic Escherichia coli O157:H7.
A:Reference number: A85480; MUID:21074935; PMID:11206551
A:Accession: E86130
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-486 <STO>
A:Cross-references: GB:AE005174; NID:q12519333; PIDN:AAGS9505.1; GSPDB:GN00145; UWGP:
A:Experimental source: strain O157:H7, substrain EDL933
C:Genetics:
A:Gene: uxuB
C:Superfamily: conserved hypothetical protein YEL070w
Query Match 30.8%; Score 775.5; DB 2; Length 486;
Best Local Similarity 36.4%; Pred. No. 6.8e-47;
Matches 177; Conservative 87; Mismatches 205; Indels 17; Gaps 9;
Qy 9 SLPANVOAPPYDIDGKPGIVHFGVGNFRAHEAFYVEQILEHA-PDWAIVGVGL-TGSD 66
Db 8 NLP--VARPSWDHRSRLESRIVHLGCGAFHRAHQALYTHLLESTDSMGICEVNLMPGND 65
Qy 67 RSKKKAEEFKAQDCLYSLTETAPSGKSTVRVMGALROYLLAPADP-EAVLKLHLYDPATRI 125
Db 66 R-VLIENLKKQQLLYTVAERG-AESTELKIIGSMKEALHPEIDGCEGILNAMARPTAI 122
Qy 126 VSMITTEGGYNINETTGAFLDLENAAVKADLNKPKPSTVFGVYVEALRRRWADGAKFTV 185
Db 123 VSLTVTEKGYCADAAAGOLDLNNPLIKHDLNPTAPSAIGYIVEALRLRREKGLKFTV 182
Qy 186 MSCDNLRHNGVARKAFGLGAKARDELAKWIEENATFPNGMVDRIPTVSAETAKLNA 245
Db 183 MSCDNLRHNGVARKAFGLGAKARDELAKWIEENATFPNGMVDRIPTVSAETAKLNA 242
Qy 246 ASGLDDDLPLVAEDFHQWLEDDQADGRPPLEKAGVQMGVDVTWYVVKIRMLNAGHYML 305
Db 243 QLVGVDPCAIAACEFROWIIEEDFVNGRPNWDKVKCAQFVADVVPPEMKLRMLNGSHSFL 302
Qy 306 CFPGLVGYENVDDAIEDSELLGNLKNVLDKVIPTTLKAPSGMTLEGYRDSVISRFSNKA 365
Db 303 AYLGVLGGYETIADTMTNPDYRKAFAFALMMQEQAPTLMSPEGTDLNAYATLLIERFSNPS 362
Qy 366 MSDQTLRIASDGCCKVQVFWETVTRRAIEDKRLSRIAFAFGIASYLEMLRGRDEKGGTYES 425
Db 363 LRHRTWQIAMDGSOKLPORLLDPVRLHLONGGSRHRLALGVAGWMRYTQGVDEQGNADIV 422

Db 361 HRTWOIAMDGSKLPQRMDSVRHMLAHDSKFDL--LALGVAGWMRYVGGVDEQGNPIEI 418
Qy 426 SEP-----TYGDAEWKLAKADDFESSKLPAFDGWRDLDTSELDOKVIVLRKIIREK 477
Db 419 SDPLLVPVQKAVQSSAEKAR-----VQSLLAIKAFGDDLPDNSLFTARVVTETLSLLAH 474
Qy 478 GVKAAT 483
Db 475 GAKATV 480
RESULT 8
G90897
probable oxidoreductase ECS2151 [imported] - Escherichia coli (strain O157:H7, substrain
C:Species: Escherichia coli
C:Date: 18-Jul-2001 #sequence_revision 18-Jul-2001 #text_change 03-Aug-2001
C:Accession: G90897
R:Hayashi, T.; Makino, K.; Ohnishi, M.; Kurokawa, K.; Ishii, K.; Yokoyama, K.; Han, C.G.
gasawara, N.; Yasunaga, T.; Kuhara, S.; Shiba, T.; Hattori, M.; Shinagawa, H.
DNA Res. 8, 11-22, 2001
A:Title: Complete genome sequence of enterohemorrhagic Escherichia coli O157:H7 and gen
A:Reference number: A99629; MUID:21156231; PMID:11258796
A:Accession: G90897
A>Status: preliminary
A:Molecule type: DNA
A:Residues: 1-486 <HAY>
A:Cross-references: GB:BA000007; PIDN:BA35574.1; PID:g13361617; GSPDB:GN00154
A:Experimental source: strain O157:H7, substrain RMD 0509952
C:Genetics:
A:Gene: ECS2151
C:Superfamily: conserved hypothetical protein YEL070w
Query Match 30.6%; Score 769.5; DB 2; Length 486;
Best Local Similarity 37.9%; Pred. No. 1.8e-46;
Matches 184; Conservative 78; Mismatches 201; Indels 23; Gaps 9;
Qy 10 LPANVQAPPYDIDGKPGIVHFGVGNFRAHEAFYVEOI--LEHAPDMAIVGVGTGSDRS 68
Db 6 LSATATLPVYDRNNLAPRIVHGFAGPHRAHQGVYADILATEHFSDMGYEVEVNLIGGE-- 63
Qy 69 KKAEFEKADCLYSLTETAPSGKSTVRMGALRDYLLAPADP--EAVLKLHVDPAIRVTS 127
Db 64 -QQTADLHQQNLTVAEMS--ADAWTARVGVVKALHVOIDGLTVLAAMECFQIAIVS 121
Qy 128 MTITEGYNINETTGAFDLENAAVKADLNKPEKSTVFGYVVEALRRRWDAAGKRAFTVMS 187
Db 122 LTITEKGFYHSPATGQLMDHPMVAADVQNPQKPTATGTVIVEALARKAAGLPAFTVMS 181
Qy 188 CDNLRHNGNVARKAFGLYAKARDPELAKWIEENATFPNGWVDRTPTVSABIAKKLNAAS 247
Db 182 CDNMPENGHVNRDVVTSYAQAVDVKLQAWIEDNVTFFSTMYVDRIVPAVTEDTLAKIEOLT 241
Qy 248 GLDDDLPLVAEDFHQWLEDOFADGRPPLEKAGVQMGVDVTDWEVVKIRMLNAGHVMCLF 307
Db 242 GVRDPAGVACPEPPQWVEDNFVAGRPWEKAGAEVSDVLPYEEMKRLMUNGSHSELAY 301
Qy 308 PGILVGENVDDAIEDSELLGNLKNYLNKQVPTLKPAGSGMTLEGYRDSVISRFSNKAAS 367
Db 302 LGYLAGYQHINCDMEDEHYRHAAYALMQEQAPTLLV--QGVVDLDYANRLIARYSNPALR 360
Qy 368 DOTLRASDGSKVQVFTWTVR--RATEDRDLRSIAFGIASYLEMLRGDRDEKGGYSES 425
Db 361 HRTWOIAMDGSKLPQRMDSVRHMLAHDSKFDL--LALGVAGWMRYVGGVDEQGNPIEI 418
Qy 426 SEP-----TYGDAEWKLAKADDFESSKLPAFDGWRDLDTSELDOKVIVLRKIIREK 477
Db 419 SDPLLVPVQKAVQSSAEKAR-----VQSLLAIKAFGDDLPDNSLFTAKVTEAYLSLLAH 474
Qy 478 GVKAAT 483
Db 475 GAKATV 480

RESULT 9
T03548
mannitol 2-dehydrogenase - Rhodobacter capsulatus
C:Species: Rhodobacter capsulatus
C:Date: 24-Mar-1999 #sequence_revision 24-Mar-1999 #text_change 08-Oct-1999
C:Accession: T03548
R:Vlcek, C.; Paces, V.; Maltsev, N.; Paces, J.; Haselkorn, R.; Fonstein, M.
Proc. Natl. Acad. Sci. U.S.A. 94, 9384-9388, 1997
A:Title: Sequence of a 189-kb segment of the chromsome of Rhodobacter capsulatus SB1
A:Reference number: Z14955; MUID:97404404; PMID:9256491
A:Accession: T03548
A>Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-478 <VLC>
A:Cross-references: EMBL:AF010496; NID:g3128256; PIDN:AAC16201.1; PID:g3128349
C:Genetics:
A:Map position: 1
C:Superfamily: conserved hypothetical protein YEL070w
Query Match 30.5%; Score 768; DB 2; Length 478;
Best Local Similarity 36.4%; Pred. No. 2.2e-46;
Matches 176; Conservative 84; Mismatches 202; Indels 22; Gaps 9;
Qy 10 LPANVQAPPYDIDGKPGIVHFGVGNFRAHEAFYVEOI--LEHAPDMAIVGVGTGSDR 67
Db 1 MPVPVAAAPRYDRAQLTPTGIVHIGLGNFRAHQAVYLDLDFGLGEGHDWALLIGAGVRAPDA 60
Qy 68 SKKAEFEKADCLYSLTETAPSGKSTVRMGALRDYLLAPADPEAVLKLHVDPAIRVTS 127
Db 61 AMR--EVLLAODGLSSVIELAP--GAHSARVIGAMIDFVAVQPGNAALIAATWARPEIRIVS 117
Qy 128 MTITEGYNINETTGAFDLENAAVKADLNKPEKSTVFGYVVEALRRRWDAAGKRAFTVMS 187
Db 118 LTVTEGYYIDPNTGTSPPDPIQADAARDPADPTAFGAIIAALKWERRAAGVAPFTVMC 177
Qy 188 CDNLRHNGNVARKAFGLYAKARDPELAKWIEENATFPNGWVDRTPTVSABIAKKLNAAS 247
Db 178 CDNPHNGHVTRDAVAVGLARLFDLAEADWIHASVAFNSMVDRTTP--ATGDRERAMAADL 236
Qy 248 GLDDDLPLVAEDFHQWLEDOFADGRPPLEKAGVQMGVDVTDWEVVKIRMLNAGHVMCLF 307
Db 237 GLDDAAPVCTCTFFQWVIEDKFPAGRPALAEKVGVTFTPHVDKFTMKIRILNGHAIAY 296
Qy 308 PGILVGENVDDAIEDSELLGNLKNYLNKQVPTLKPAGSGMTLEGYRDSVISRFSNKAAS 367
Db 297 PGLADIATFVHEAMADPQIRAFNLRLTAEVLPVPPVPGDLDNDYKALIIERFSNPEVA 356
Qy 368 DOTLRASDGSKVQVFTWTVRRAIEDKRLSRIAFGIASYLEMLRGDRDEKGGYSES 427
Db 357 DTVRLCLDGSNRQPKFTIPSIRDIRIATGGSFDFGLVLLSALWCRCYCFGTTESGAELAFND 416
Qy 428 PTYGDAAEWKLAKADDFESSKLPAFDGWRDLDT--SELDOKVIVLR-----KIIREKV 479
Db 417 PNW--ERLVAVA---QAQDRPA--AWLEMEDIAEVRNPVAVTSTTTALKALWARGS 468
Qy 480 KAAI 483
Db 469 RAVI 472
RESULT 10
H85719
probable oxidoreductase ydfi [imported] - Escherichia coli (strain O157:H7, substrain
C:Species: Escherichia coli
C:Date: 16-Feb-2001 #sequence_revision 16-Feb-2001 #text_change 14-Sep-2001
C:Accession: H85719
R:Perna, N.T.; Plunkett III, G.; Burland, V.; Mau, B.; Glasner, J.D.; Rose, D.J.; May
iller, L.; Grotbeck, E.J.; Davis, N.W.; Lim, A.; Dimantanta, E.; Potamouis, K.; Apoda
Nature 409, 529-533, 2001
A:Title: Genome sequence of enterohemorrhagic Escherichia coli O157:H7.
A:Reference number: A85480; MUID:21074935; PMID:11206551
A:Accession: H85719
A>Status: preliminary

A:Molecule type: DNA
A:Residues: 1-486 <STO>
A:Cross-references: GB:AF005174; NID:g12515111; PIDN:AAG56220.1; GSPDB:GN00145; UWGP:Z21
A:Experimental source: strain 0157:H7, substrain EDL933
C:Genetics:
C:Gene: ydfI
C:Superfamily: conserved hypothetical protein YEL070w

Query Match 30.3%; Score 763.5; DB 2; Length 486;
Best Local Similarity 37.7%; Pred. No. 4.8e-46;
Matches 183; Conservative 78; Mismatches 202; Indels 23; Gaps 9;

QY 10 LPANVQAPPYDIDGKPGIVHFGVGNFFRAHEAFVEQI-LEHAPDMAIVGVLGTGSDRS 68
DB 6 LSAKATLPVYDRNRLAPRIVHLGFGAPRAHQGVYADILATXHFSDWCYVEVNLIGGE-- 63
QY 69 KKAEEFKAQDCLSLTETAPSGKSTVRVMGALRDYLLAPADP-EAVLKLHVDPAIRVS 127
DB 64 -QQIADLHQDNLTYVAEMS-ADAWTARVGVVKAALHVQDGLERVLAAACEQIAIVS 121
QY 128 MTITEGGYNINETTGAFOLENAAVKADLNKPEKSTVFGVYVEALRRWDAGGKRAFTVMS 187
DB 122 LTITEKGFHFSPATGQLMDDHPWAAVQNPQKATGIVEALARRKAAGLPAFTVMS 181
QY 188 CDNLRHNGNVARKAFGLYAKARDELAKWIEENATFPNGWVDRIPTTYSABIAKLNAAAS 247
DB 182 CDNMPENGHVMDVVTSAQAVDVKLAQWIEDNVTFPSTWMDRIVPAVTEDTLAKIEQLT 241
QY 248 GLDDDLPLVAEDFHQWLEDOFADGRPPLEKAGVQMGVDVTDWEYKIRMLNAGHVMLCF 307
DB 242 GVRDPAGVACEPFQWVIEDTFVAGRPQWENAGATLVADVVPFEEMKLRMLNGSHSFLAY 301
QY 308 PGILVYENVDDAIEDSELLGNLKNYLNKDVPTLKAPSGMTLEGYRDSVTSRFSNKAMS 367
DB 182 CDNMPENGHVMDVVTSAQAVDVKLAQWIEDNVTFPSTWMDRIVPAVTEDTLAKIEQLT 241
QY 248 GLDDDLPLVAEDFHQWLEDOFADGRPPLEKAGVQMGVDVTDWEYKIRMLNAGHVMLCF 307
DB 242 GVRDPAGVACEPFQWVIEDTFVAGRPQWENAGATLVADVVPFEEMKLRMLNGSHSFLAY 301
QY 308 PGILVYENVDDAIEDSELLGNLKNYLNKDVPTLKAPSGMTLEGYRDSVTSRFSNKAMS 367
DB 302 LGYLAGYQHINDCDDDDNYRLTAQALMLREQATLVK-OGVDLQRYADQLIARYNPALR 360
QY 368 DOTLRASDGGSKVQVFTETVR-RAIEDKRLSRIATGASLYLEMLRGRDEKGGTYESSE 427
DB 242 GVRDPAGVACEPFQWVIEDTFVAGRPQWENAGATLVADVVPFEEMKLRMLNGSHSFLAY 301
QY 308 PGILVYENVDDAIEDSELLGNLKNYLNKDVPTLKAPSGMTLEGYRDSVTSRFSNKAMS 367
DB 302 LGYLAGYQHINDCDDDDNYRLTAQALMLREQATLVK-OGVDLQRYADQLIARYNPALR 360
QY 426 SEP-----TYGDAEWKLAADDFESSLKLPAFDGWRDLDTSELDQKIVLVRLKIREK 477
DB 419 SDPLLPVIOKAVQSSAEGKAR-----VQSLLAIKAFGGDLDPNLSLFTAKVTEAYLSLLAH 474
QY 478 GVKAAT 483
DB 475 GAKATV 480

RESULT 11
AD0679
fructuronate reductase (EC 1.1.1.57) - Salmonella enterica subsp. enterica serovar Typhi
C:Species: Salmonella enterica subsp. enterica serovar Typhi
A:Note: this species has also been called Salmonella typhi
C:Accession: AD0679
C:Date: 09-Nov-2001 #sequence_revision 09-Nov-2001 #text_change 27-Nov-2001
R:Parkhill, J.; Dougan, G.; James, K.D.; Thomson, N.R.; Pickard, D.; Wain, J.; Churcher,
th, T.; Connor, P.; Cronin, A.; Davis, P.; Davies, R.M.; Dowd, L.; White, N.; Farrar,
S.; Moule, S.; O'Gaora, P.
Nature 413, 848-852, 2001
A:Authors: Parry, C.; Quail, M.; Rutherford, K.; Simmonds, M.; Skelton, J.; Stevens, K.;
A:Title: Complete genome sequence of a multiple drug resistant Salmonella enterica serov
A:Reference number: AB0502; PMID:11677608
A:Accession: AD0679
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-488 <PAR>
A:Cross-references: GB:AL513382; PIDN:CAD01805.1; PID:g16502650; GSPDB:GN00176
C:Genetics:
C:Gene: STY1553
C:Superfamily: conserved hypothetical protein YEL070w
C:Keywords: oxidoreductase

Query Match 30.2%; Score 759.5; DB 2; Length 488;
Best Local Similarity 36.3%; Pred. No. 9.2e-46;
Matches 175; Conservative 87; Mismatches 205; Indels 15; Gaps 7;

QY 10 LPANVQAPPYDIDGKPGIVHFGVGNFFRAHEAFVEQI-LEHAPDMAIVGVLGTGSDRS 68
DB 6 LSAKATLPYDRALAAARMVHLGFGAFRAHQGVYITDLAAEQHSDWGYVEVNLIGGE-- 63
QY 69 KKAEEFKAQDCLSLTETAPSGKSTVRVMGALRDYLLAPADP-EAVLKLHVDPAIRVS 127
DB 64 -QQIADLHQDNLTYVAEMS-ADAWTARVGVVKAALHVQDGLERVLAAACEQIAIVS 121
QY 128 MTITEGGYNINETTGAFOLENAAVKADLNKPEKSTVFGVYVEALRRWDAGGKRAFTVMS 187
DB 122 LTITEKGVCHSPATGQLLLEHPMIAADLQNPQPLTAPGIIIVEALARRKAAGLPAFTVMS 181
QY 188 CDNLRHNGNVARKAFGLYAKARDELAKWIEENATFPNGWVDRIPTTYSABIAKLNAAAS 247
DB 182 CDNMPENGHVTRQVVTAYAREVDAELAIWIBQNTFFPSTWMDRIVPAVTPETLTKIEQLT 241
QY 248 GLDDDLPLVAEDFHQWLEDOFADGRPPLEKAGVQMGVDVTDWEYKIRMLNAGHVMLCF 307
DB 242 GVRDPAGVACEPFQWVIEDTFVAGRPQWENAGATLVADVVPFEEMKLRMLNGSHSFLAY 301
QY 308 PGILVYENVDDAIEDSELLGNLKNYLNKDVPTLKAPSGMTLEGYRDSVTSRFSNKAMS 367
DB 302 LGYLAGYQHINDCDDDDNYRLTAQALMLREQATLVK-OGVDLQRYADQLIARYNPALR 360
QY 368 DOTLRASDGGSKVQVFTETVR-RAIEDKRLSRIATGASLYLEMLRGRDEKGGTYESSE 427
DB 361 HHTWQIAMDGSQKLPQRMDSVRWHLANHSDFLLALGVAGWRYVGVGDEQKRAIDVSD 420
QY 428 PTYGDAEWKLAADDFESSLKLPAFDGWRDLDTSELDQKIVLVRLKIREK 481
DB 421 PLLPVIQAVANSEBEGASRVK--ALLGMAEIFGNDLPQARFTQKQVQAYDSLITYGAKA 478
QY 482 AI 483
DB 479 SV 480

RESULT 12
AF0156
probable D-mannanase oxidoreductase YP01280 [imported] - Yersinia pestis (strain CO92
C:Species: Yersinia pestis
C:Date: 02-Nov-2001 #sequence_revision 02-Nov-2001 #text_change 09-Nov-2001
C:Accession: AF0156
R:Parkhill, J.; Wren, B.W.; Thomson, N.R.; Titball, R.W.; Holden, M.T.G.; Prentice, M
deno-Tarraga, A.M.; Chillingworth, T.; Cronin, A.; Davies, R.M.; Davis, P.; Dougan, G
ll, M.; Rutherford, K.; Simmonds, M.; Skelton, J.; Stevens, K.; Whitehead, S.; Barrel
Nature 413, 523-527, 2001
A:Title: Genome sequence of Yersinia pestis, the causative agent of plague.
A:Reference number: AB0001; PMID:21470413; PMID:11586360
A:Accession: AF0156
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-490 <KUR>
A:Cross-references: GB:AL50842; PIDN:CAC90113.1; PID:g15979333; GSPDB:GN00175
C:Genetics:
C:Gene: YP01280
C:Superfamily: conserved hypothetical protein YEL070w

Query Match 29.9%; Score 753.5; DB 2; Length 490;
Best Local Similarity 37.8%; Pred. No. 2.4e-45;
Matches 171; Conservative 80; Mismatches 190; Indels 11; Gaps 6;

QY 9 SLPANVQAPPYDIDGKPGIVHFGVGNFFRAHEAFVEQI-LEHAPDMAIVGVLGTGSDR 67
DB 8 ALPSTVQPNYDRALRSRIVHIFGAFRAHQALLTYRVLNTGGDWGICEVSLFGAD- 66
QY 68 SKKAAEFKAQDCLSLTETAPSGKSTVRVMGALRDYLLAPADP-EAVLKLHVDPAIRV 126
DB 67 --TLIQELRKODHLEFSVLEKGAQGNQAI-VVGSVCESVHARLDGIMQVLAKLVEPQVAIV 123

Qy	127	SMTTEGGYNINETTGAFDLENAAVKADLNKPEKPSVFGYVVEALRRRRWDAGKCAETVM	186
		: : : : : : : : : : : : :	
Db	124	SLATTEKGYCEPTQGLDQNESIRADLAVPNAPTAPGYLVEALRLRLGLPPETVL	183
		: : : : : : : : : : : : :	
Qy	187	SCDMLRINGNVARKAFLGYAKARPELAKMTEENATPNGMVDRIITPTVSAETAKKLNAA	246
		: : : : : : : : : : : : :	
Db	184	SCDNI PENGDVVRNAVGLATVRDPALANIQIVTPTNTMVDRIVPAATAETLQETADT	243
		: : : : : : : : : : : : :	
Qy	247	SGLDDDLPLVAEDPHOWLEDOFADGRPPLEKAGQVMQVGDVTDWYEVKIRMLNAGHYMLC	306
		: : : : : : : : : : : : :	
Db	244	LGVADSCGIACEPFIQMWVEDKFVAGRPDQCVAGVQLVDVFLPPEEMKLRMLNGSHSYLS	303
		: : : : : : : : : : : : :	
Qy	307	FPGLVLGYENVDDAIEDSELLGNLKNYLNKDVIPTLKAPSGMTLEGYRDSVISRFSNKAM	366
		: : : : : : : : : : : : :	
Db	304	YLGVLGYQYHNDQCMADENRYLTARRLMNEQAPTLKV-TGIDILNAYADOLIERYCNPAL	362
		: : : : : : : : : : : : :	
Qy	367	SDQTLRTASDGCXKVQVFWETVTRARATEDKRLDSRIAFGIASYLEMLRGRDEKGGTVESS	426
		: : : : : : : : : : : : :	
Db	363	KHRTWQIAMDGSQKLPQRMLDXSVRWHLKQGNAYPCLALGIAGWMRYVGGIDNDGQVIDIR	422
		: : : : : : : : : : : : :	
Qy	427	EPTYGDRAEWKLAKADD---FESSKLPAFDG	454
		: : : : : : : : : : : : :	
Db	423	DPMWDSFKQCVAAESDGAARVQSILLTKALFG	454

RESULT 13
H91011
probable oxidoreductase EC33064 [imported] - Escherichia coli (strain O157:H7, substrain
C;Species: Escherichia coli
C;Date: 18-Jul-2001 #sequence_revision 18-Jul-2001 #text_change 03-Aug-2001
C;Accession: H91011
R:Hayashi, T.; Makino, K.; Ohnishi, M.; Kurokawa, K.; Ishii, K.; Yokoyama, K.; Han, C.G.
gasawara, N.; Yasunaga, T.; Kuhara, S.; Shiba, T.; Hattori, M.; Shinagawa, H.
DNA Res. 8, 11-22, 2001
A:Title: Complete genome sequence of enterohemorrhagic Escherichia coli O157:H7 and gen
A:Reference number: A99629; MUID:21156231; PMID:11258796
A:Accession: H91011
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-488 <HAY>
A:Cross-references: PIDN:BA000007; PIDN:BA036487.1; PID:gl33362533; GSPDB:GN00154
A:Experimental source: strain O157:H7, substrain RMD 0509952
C:Genetics:
A:Gene: ECs3064
C:Superfamily: conserved hypothetical protein YEL070w

Query Match	29.7%;	Score	746.5;	DB 2;	Length	488;			
Best Local Similarity	37.7%;	Pred. No.	7.6e-45;						
Matches	159;	Conservative	81;	Mismatches	175;	Indels	7;	Gaps	5;
Qy	9	SLPANQVAPPYDIDGIRKPGIVHFGVGNFFRAHEAFYVEQILE-HAPDWAIVGVGLTGSDR	67						
Db	8	TLPHVHAPRYDROQLQSRIVHFGFAGHRAHQALLTDRLVNAQGGDWGMCIEISLFSGD-	66						
Qy	68	SKKKAEEFKADCLYSLTEAPSKSTVRYMGALRDYLLAPADP-EAVLKLHVPDAIRIV	126						
Db	67	--QUMSOLRAQNHILYVLEKGADGNQAI-IVGAVHECHNAKLDSLAATIERCFEPQVAIV	123						
Qy	127	SMTITEGYINETTGAFDLENAAVKADLKPKPSTVFGVVEALRRRDAGCKAFTVM	186						
Db	124	SLTITEGYCIDPATGALDTSNPRIIHDLOTPPEPHSAPGILVEALKRRRRRGTLTPFTVL	183						
Qy	187	SCDLNRHNGVNRKAFGLYAKARDPELAKWTEENATFPGMVDRITPVSAEIAKKNAA	246						
Db	184	SCDNI PDNGHVHVNKAVLGMAEKRSPELAGWIKHEVSPGTVDRIVPAAATDESIAETISQH	243						
Qy	247	SGLDDDLPLVAEDPHQWLEQFADGRPPLEKAGVQWGDYTDWEYKIRMLNAGHVMLC	306						
Db	244	LGVNDPCAISCEPIQWVVEDFVAGRPWEVAGVQWVNDVLPWEEKMLRMINGSHSFLA	303						
Qy	307	FPGLVGYENVDDAIEDSELLGNLKNYLKNDKVIPTLKAPSGMTLEGYRDSVISRFSNKM	366						

Db 304 YLGYLSGFAMHSDCMQDRAPFAHAARTMLMDQAFTLR - KDVDLTQTVAADKLIARFANPAL 362

Qy 367 SDQTLRIASDGCXKVQVFWTETVRRAIEDKRDLRSIAFGIASYLEMLRGRDEKGCTYESS 426
: : ||| : : : : | : | : | : | : | :

Db 363 KHKTWQIAMDGSOKLPQRLMAGIRIHLEGRETDSLLALGVAGWMRYVSGVDDACNAIDVR 422

Qy 427 EP 428
: :

Db 423 DP 424
: :

RESULT 14
B85856
Probable oxidoreductase yelQ [imported] - Escherichia coli (strain O157:H7, substrain C; Species: Escherichia coli)
C; Date: 16-Feb-2001 #sequence_revision 16-Feb-2001 #text_change 14-Sep-2001
C; Accession: B85856
R.; Perna, N.T.; Plunkett III, G.; Burland, V.; Mau, B.; Glasner, J.D.; Rose, D.J.; May-
iller, L.; Grobeck, E.J.; Davis, N.W.; Lim, A.; Dimalanta, E.; Potamousis, K.; Apoda-
Nature 409, 529-533, 2001
A; Title: Genome sequence of enterohemorrhagic Escherichia coli O157:H7.
A; Reference number: B85480; MUID:21074935; PMID:11206551
A; Accession: B85856
A; Status: preliminary
A; Molecule type: DNA
A; Residues: 1-488 <STO>
A; Cross-references: GB:AE005174; NID:g12516498; PIDN:AAG57310.1; GSPDB:GN00145; UWGP:
A; Experimental source: strain O157:H7, substrain EDL933
C; Genetics:
A; Gene: yelQ
C; Superfamily: conserved hypothetical protein YEL070w

[illegible]

Search completed: May 8, 2003, 16:34:29
Job time : 47 secs

GenCore version 5.1.4_p5_4578
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: May 8, 2003, 16:25:56 ; Search time 27 Seconds
(without alignments)
745.038 Million cell updates/sec

Title: US-09-926-163B-2

Perfect score: 2516

Sequence: 1 MITRETLKSLPANVQAPPD.....KVIVLRKIIRKGVKAAIPA 485

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 112892 seqs, 41476328 residues

Total number of hits satisfying chosen parameters: 112892

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : SwissProt_40:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query	Match	Length	ID	Description
1	789	31.4	502	1	YE10_YEAST	P39941 saccharomyc
2	774.5	30.8	486	1	UXUB_ECOLI	P39160 escherichia
3	771.5	30.7	486	1	YDFI_ECOLI	P77260 escherichia
4	739.5	29.4	488	1	YE10_ECOLI	P33029 escherichia
5	725.5	28.8	476	1	MTLK_RHOSH	P33216 rhodobacter
6	583	23.2	465	1	DALD_RALSO	P58708 ralstonia s
7	569	22.6	463	1	DALD_VERPE	P58709 versinia pe
8	534	21.2	455	1	DALD_KLEPN	O52720 klebsiella
9	354	14.1	228	1	YTUB_ERWHE	Q47826 erwinnia her
10	285	11.3	483	1	UXAB_ECOLI	P24214 escherichia
11	258	10.3	50	1	POR_GLUOX	P80354 gluconobact
12	249.5	9.9	384	1	MTLD_CLOAB	O65992 clostridium
13	205	8.1	382	1	MTLD_ECOLI	P09424 escherichia
14	199	7.9	382	1	MTLD_BACST	Q5421 bacillus st
15	193	7.7	357	1	MTLD_ENTFA	P27543 enterococcu
16	192	7.6	382	1	MTLD_STRMU	Q02418 streptococc
17	190	7.6	366	1	MTLD_BACSU	P42957 bacillus su
18	184.5	7.3	374	1	MTLD_BACHD	Q9k681 bacillus ha
19	184	7.3	382	1	MTLD_KLEPN	Q9xbm6 klebsiella
20	138.5	5.5	385	1	MTLD_BUCAL	P57634 buchnera ap
21	123.5	4.9	691	1	EFGB_BACSU	P80868 bacillus su
22	119	4.7	3685	1	DMD_HUMAN	P11532 homo sapien
23	118	4.7	3678	1	DMD_MOUSE	P11531 mus musculu
24	116.5	4.6	683	1	RPC_BPPHC	P08979 bacterioph
25	114	4.5	754	1	RPLI_INCUJ	P19703 influenza c
26	110	4.4	692	1	EFGB_BACHD	Q92917 bacillus ha
27	109.5	4.4	3680	1	DMD_CANFA	O97592 canis fami
28	107	4.3	1703	1	SNF2_YEAST	P22082 saccharomyc
29	104.5	4.2	619	1	HSCA_HAEIN	P44669 haemophilus
30	103.5	4.1	775	1	YTXL_XENLA	P14380 xenopus lae
31	103	4.1	372	1	FTZL_PYRAB	Q9v280 pyrococcus
32	103	4.1	675	1	VP55_YEAST	Q92331 saccharomyc
33	103	4.1	692	1	EFGB_MYCPU	Q98qd8 mycoplasma

RESULT 1

ID	YE10_YEAST	-STANDARD:	PRT:	502 AA.
AC	P39941;			
DT	01-FEB-1995 (Rel. 31, Created)			
DT	01-FEB-1995 (Rel. 31, Last sequence update)			
DT	16-OCT-2001 (Rel. 40, Last annotation update)			
DE	Hypothetical 56.5 kDa protein in HXT8 5' region and in HXT17-COS10 intergenic region.			
DE	YE1070W AND (YNR073C OR N3810).			
OS	Saccharomyces cerevisiae (Baker's yeast).			
OC	Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;			
OC	Saccharomycetales; Saccharomycetaceae; Saccharomycetes.			
OX	NCBI_TaxID=4932;			
RN	[1]			
RP	SEQUENCE FROM N.A. (YE1070W).			
RC	STRAIN=S288c / AB972;			
RA	Dietrich F.S., Mulligan J.T., Hennessey K.M., Allen E., Araujo R.,			
RA	Aviles E., Berno A., Brennan T., Carpenter J., Chen E., Cherry J.M.,			
RA	Chung E., Duncan M., Guzman E., Hartzell G., Hunnicke-Smith S.,			
RA	Hyman R., Kayser A., Komp C., Lashkari D., Lew H., Lin D.,			
RA	Mosedale D., Nakahara K., Namath A., Norgren R., Oefner P., Oh C.,			
RA	Petel F.X., Roberts D., Sehl P., Schramm S., Shogren T., Smith V.,			
RA	Taylor P., Wei Y., Yelton M., Botstein D., Davis R.W.;			
RA	Submitted (DEC-1994) to the EMBL/GenBank/DBJ databases.			
[2]	SEQUENCE FROM N.A. (YNR073C).			
RL	Andre B., Iraqi Houssaini I., Urrestarazu L.A., Vissers S.;			
RN	Submitted (MAY-1996) to the EMBL/GenBank/DBJ databases.			
[3]	SEQUENCE OF 1-439 FROM N.A. (YNR073C).			
RP	STRAIN=S288c;			
RC	MEDLINE=97060021; PubMed=8904342;			
RX	Levesque H., Nicaud J.-M., Lepingle A., Gaillardin C.;			
RT	"Sequencing of a 9.2 kb telomeric fragment from the right arm of			
RT	Saccharomyces cerevisiae chromosome XIV.";			
RL	Yeast 12:289-295(1996).			
CC	!- SIMILARITY: BELONGS TO THE MANNITOL DEHYDROGENASES FAMILY.			
CC	-----			
CC	This SWISS-PROT entry is copyright. It is produced through a collaboration			
CC	between the Swiss Institute of Bioinformatics and the EMBL outstation -			
CC	the European Bioinformatics Institute. There are no restrictions on its			
CC	use by non-profit institutions as long as its content is in no way			
CC	modified and this statement is not removed. Usage by and for commercial			
CC	entities requires a license agreement (see http://www.isb-sib.ch/announce/			
CC	or send an email to license@isb-sib.ch).			
CC	-----			
DR	EMBL; U18795; AAB65017.1; -			
DR	EMBL; Z71888; CAA96356.1; -			
DR	EMBL; Z71689; CAA96358.1; -			
DR	EMBL; X86790; CAA60486.1; -			
DR	SGD; S0000796; YEL070W.			
DR	SGD; S0005356; YNR073C.			
DR	InterPro; IPR0000669; Mannitol_dh.			
DR	Pfam; PF01232; Mannitol_dh; 1.			
DR	PRINTS; PR00084; MTLHDHGNASE.			

```
DR PROSITE; PS00974; MANNITOL_DHGENASE; 1.
KW Hypothetical protein; Oxidoreductase; NAD.
SQ SEQUENCE 502 AA; 56470 MW; A7E0CC01AADC1B2A CRC64;

Query Match 31.4%; Score 789; DB 1; Length 502;
Best Local Similarity 35.7%; Pred. No. 1.3e-46;
Matches 179; Conservative 90; Mismatches 195; Indels 38; Gaps 10;

QY 2 ITRRLKSLPANVQAPPYDIDGKPGIVHFGVGNFRAHEAFVVEOILEH-PDMAIVGVL-TGSD 59
DB 12 LNAKTLKSFESTLPIPTYPREGVKQGVHLGVGAFHRSHLAVFMHRLMOEHKLDKWSICG 71
QY 60 VGLTGSRRKKAEEFKAQDCLYSTETAPSGKSTVRVMGALRDYLLAPADPEAVLKHIV 119
DB 72 VGLMKADLMRDA--MKAQDCLYTLVERGIK-DTNAYIVGSIYAYMYAPDDPRAVIEKWA 128
QY 120 DPAIRIVSMITTEGGYNINETGAFDLENAAVKADLNKPEKSTVGYVVEALRRRWADAG 179
DB 129 NPDTHIVSLTVTENGYYHSEATNSLMTDAPEIINDLNHPKPDTLGYLYEALLRYKRG 188
QY 180 GRAFTVMSCDNLRHNGNVARKAFGLYAK-ARPELAKWIEENATFPNGMVDRTTPTVSAE 238
DB 189 LPFTTMSCDNPNQGVTVKTMVLVAFKLLKDEKFAWIEDKVTSPNSMVDRTVPRCTDK 248
QY 239 IAKLNAASGLDDDLPLVAEDFHQWLEDFADGRPPLEKAGVQMVGDVTDWEYVKIRML 298
DB 249 ERKYVADTWGIXDQCPVAEPIQWLEDFNSDGRPPWELVGVQVVKDVSVELMKLRLL 308
QY 299 NAGHVMLCFPGILGVENVDDAIEDSELGNLKNYLNKDVITPLKAPSGMTLEGYRDSVI 358
DB 309 NGHGSAMGVLGYLAGTYIHEVNDPTINKYIRVLMREEVILPLPKVPGVDFEETASVL 368
QY 359 SPSNKAMSDQTLRIASDGCCKVQVFWTETVRRRAIEDKRD--LSRTAFGIASVLEMLRGR 416
DB 369 ERESNPAIQDTVARICLMSGKMPKVLPSIYEQLR-KPDGKYKLLAVCVAGHFRYLTV 427
QY 417 DEKGGTYESEPTYGDAEWKLAKADDFESSKLPLAPFDGWRD-----LDTSEL-DOK 466
DB 428 DMNGKPFIEDP-----MAPTLKAAAVKGGKDPHELLNIEVLFSPERDKN 473
QY 467 VIVLR-----KIIREKGVKAAI 483
DB 474 EFVAQLTHSLETYDKGPIAAI 495

RESULT 2
UXUB_ECOLI STANDARD; PRT; 486 AA.
AC P39160;
DT 01-FEB-1995 (Rel. 31, Created)
DT 01-FEB-1995 (Rel. 31, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE D-mannonate oxidoreductase (EC 1.1.1.57) (Fructuronate reductase).
GN UXUB OR B4323
OS Escherichia coli.
OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
OX Escherichia.
OC NCBI_TaxID=562;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=K12 / MG1655;
RX MEDLINE=95334362; PubMed=7610040;
RA Burland V.D., Plunkett G. III, Sofia H.J., Daniels D.L.,
RA Blattner F.R.;
RT "Analysis of the Escherichia coli genome VI: DNA sequence of the
RN Nucleic Acids Res. 23:2105-2119(1995).
RL [2]
RP SEQUENCE FROM N.A.
RC STRAIN=K12 / W3110;
RA Mizobuchi K.;
RL Submitted (SEP-1992) to the EMBL/GenBank/DBJ databases.
CC -!- CATALYTIC ACTIVITY: D-mannonate + NAD(+) = D-fructuronate + NADH.
```



```
OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
OC Escherichia.
OX NCBI_TaxID=562;
RN SEQUENCE FROM N.A.
RC STRAIN=K12 / MG1655;
RX MEDLINE=97426617; PubMed=9278503;
RA Blattner F.R., Plunkett G. III, Bloch C.A., Perna N.T., Burland V.,
RA Riley M., Collado-Vides J., Glasner J.D., Rode C.K., Mayhew G.F.,
RA Gregor J., Davis N.W., Kirkpatrick H.A., Goeden M.A., Rose D.J.,
RA Mau B., Shao Y.;
RT "The complete genome sequence of Escherichia coli K-12.";
RL Science 277:1453-1474(1997).
RN SEQUENCE FROM N.A.
RC STRAIN=K12;
RX MEDLINE=97251357; PubMed=9097039;
RA Aiba H., Baba T., Fujita K., Hayashi K., Inada T., Isono K.,
RA Kitagawa M., Kasai H., Kashimoto K., Kimura S., Kitakawa M.,
RA Motomura K., Nakade S., Nakamura Y., Nishio Y.,
RA Oshima T., Saito N., Sampei G., Seki Y., Sivasubram S.,
RA Yamamoto Y., Takeda J., Takemoto K., Takeuchi Y., Wada C.,
RA Yanaguchi Y., Horiuchi T.;
RT "A 570-kb DNA sequence of the Escherichia coli K-12 genome
corresponding to the 28.0-40.1 min region on the linkage map.";
RL DNA Res. 3:363-377(1996).
CC -!- SIMILARITY: BELONGS TO THE MANNITOL DEHYDROGENASES FAMILY.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; AE000251; AAC74615.1; -
DR EMBL; D90797; BAA15243.1; -
DR EMBL; D90798; BAA15247.1; -
DR EcoGene; EG13821; ydgi.
DR InterPro; IPR000669; Mannitol_dh.
DR Pfam; PF01232; Mannitol_dh; 1.
DR PRINTS; PR00084; MTLDHORGNASE.
DR PROSITE; PS00974; MANNITOL_DHCENASE; 1.
DR KW Hypothetical protein; Oxidoreductase; NAD; Complete proteome.
DR NP_BIND 24 35 NAD (BY SIMILARITY).
DR SEQUENCE 486 AA; 53685 MW; 96F3BF07AE114B70 CRC64;
Query Match 30.7%; Score 771.5; DB 1; Length 486;
Best Local Similarity 37.7%; Pred. No. 1.9e-45;
Matches 183; Conservative 81; Mismatches 199; Indels 23; Gaps 9;
QY 10 LPANVQAPPYDIDGKPGIVHFGVGNFRAHEAFVVEQI-LEHAPDWAIVGLTGSDBS 68
DB 6 LSAKATPLVDLNNAPRIVHFGAFHAFHOGVYADILATEHFDSDGWYVEVNLIGGE-- 63
QY 69 KKAEEFKAQCLYSITETAPSGKSTVRVMGALRDYLLAPADP-EAVLKLHLDVPAIRVS 127
DB 64 "QIADLQODNLYTVAEMS-ADWTVARVGVVKALHVDIGLETVLAACEPOIAIVS 121
QY 128 WTITEGGYNNITTCADFLENAVKAADLNKPEKPTVFYGVVVEALRRWRDAGGKAFTVMS 187
DB 122 LTITEKGYFHSPTAQMLDHPMVAADYQNPQHPKTATGVIVEALARRKAAGLPAFTVMS 181
QY 188 CDNLHNGNVARKAFGLYAKARDPLEAKWIENATFPNGWDRITPTVSAETAKLNAA 247
DB 182 CDMNPENGVHVRDVTYSQAQAVDKLAQWIEDNVTFPSTMDRIYPAVTEDTLAKIEQLT 241
QY 248 GLDDDLPLVAEDFHOWLEDOFADGRPPLEKAGVQMGVDGTDWEYVKIRMLNAGHVMLCF 307
DB 242 GVRDPAGVACEPFRWVIEDNFVAGRPWEKAGAEVSDVLPYEEMKMLNGLNGSHSFLAY 301
```

```
QY 308 PGILVGVENVDDAIEDSELLGNLKNLYNKDVITPTLKAPSGMTLEGYRDSVSRFSNKAMS 367
DB 302 LGYLAGYQHINDCMEDEHYRYAAYGLMLQEQAPTLKV-QGVLDQDYANRLIARYSNPALR 360
QY 368 DOTLRASDGSCKVQVFWTETVR--RAIEDKRDLRSIAFGIASYLEMLRGRDEKGGTYES 425
DB 361 HRTWQIAMDGSKLPQRLMDSVRWHLAHSKFDL--LALGVAGWMRYGVGVDEQGNPIEI 418
QY 426 SEP-----TYGDAEWKLAKADDFESSKLKLPADFQWRDLDTSELDQKVVILRKIREK 477
DB 419 SDPLLPVIOKAVOSSAEGKAR----VQSLLAIFAIFGDDLPDNLSTFARVTETLYLSLLAH 474
QY 478 GVKAAI 483
DB 475 GAKATV 480
RESULT 4
YEIQ_ECOLI
ID YEIQ_ECOLI STANDARD; PRT; 488 AA.
AC P33029; P94760;
DT 01-FEB-1994 (Rel. 28, Created)
DT 01-FEB-1994 (Rel. 28, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Hypothetical oxidoreductase yeiQ (EC 1.-.-.-).
GN YEIQ OR B2172.
OS Escherichia coli.
OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
OC Escherichia.
OX NCBI_TaxID=562;
RN SEQUENCE FROM N.A.
RC STRAIN=K12 / BHB2600;
RA Richterich P., Lakey N., Gryan G., Jaehn L., Mintz L., Robison K.,
RA Church G.M.;
RL Submitted (OCT-1993) to the EMBL/GenBank/DBJ databases.
RN SEQUENCE FROM N.A.
RC STRAIN=K12 / MG1655;
RX MEDLINE=97426617; PubMed=9278503;
RA Blattner F.R., Plunkett G. III, Bloch C.A., Perna N.T., Burland V.,
RA Riley M., Collado-Vides J., Glasner J.D., Rode C.K., Mayhew G.F.,
RA Gregor J., Davis N.W., Kirkpatrick H.A., Goeden M.A., Rose D.J.,
RA Mau B., Shao Y.;
RT "The complete genome sequence of Escherichia coli K-12.";
RL Science 277:1453-1474(1997).
RN SEQUENCE FROM N.A.
RC STRAIN=K12;
RX MEDLINE=97251358; PubMed=9097040;
RA Itoh T., Aiba H., Baba T., Fujita K., Hayashi K., Inada T., Isono K.,
RA Kasai H., Kimura S., Kitakawa M., Kitagawa M., Makino K., Miki T.,
RA Mizobuchi K., Mori H., Mori T., Motomura K., Nakade S., Nakamura Y.,
RA Nishimoto H., Nishio Y., Oshima T., Saito N., Sampei G., Seki Y.,
RA Sivasubram S., Tagami H., Takeda J., Takemoto K., Wada C.,
RA Yamamoto Y., Horiuchi T.;
RT "A 460-kb DNA sequence of the Escherichia coli K-12 genome
corresponding to the 40.1-50.0 min region on the linkage map.";
RL DNA Res. 3:379-392(1996).
CC -!- SIMILARITY: BELONGS TO THE MANNITOL DEHYDROGENASES FAMILY.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; U00007; AAA60520.1; -
DR EMBL; AE000306; AAC75233.1; -
DR EMBL; D90849; BAA15981.1; -
DR EcoGene; EG12036; yeiQ.
```



```
RC STRAIN-CM11000;
RA MEDLINE=21681879; PubMed=11823852;
RA Salanoubat M., Genin S., Artiguenave F., Gouzy J., Mangenot S.,
RA Arlat M., Billault A., Brottier P., Camus J.C., Cattolico L.,
RA Chandler M., Cholsne N., Claudel-Renard C., Cunnean S., Demange N.,
RA Gaspin C., Lavie M., Moisan A., Robert C., Saurin W., Schiex T.,
RA Siguer P., Thebaud P., Whalen M., Wincker P., Levy M.,
RA Weissenbach J., Boucher C.A.;
RT "Genome sequence of the plant pathogen Ralstonia solanacearum.";
RL Nature 415:497-502(2002).
CC -!- CATALYTIC ACTIVITY: D-arabinitol + NAD(+) = D-xylulose + NADH.
CC -!- PATHWAY: D-arabinitol catabolism; first step.
CC -!- SIMILARITY: BELONGS TO THE MANNITOL DEHYDROGENASES FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; AL646068; CAD15836.1;
DR InterPro: IPR000669; Mannitol_dh.
DR Pfam: PF01232; Mannitol_dh; 1.
DR PRINTS: PR00084; MTLDRHGNASE.
DR PROSITE: PS00974; MANNITOL_DHGENASE; FALSE_NEG.
KW Oxidoreductase; NAD; Complete proteome.
SQ SEQUENCE 465 AA; 50630 MW; 47095425D4836342 CRC64;

Query Match 23.28; Score 583; DB 1; Length 465;
Best Local Similarity 35.98; Pred. No. 1.3e-32;
Matches 142; Conservative 63; Mismatches 177; Indels 14; Gaps 8;

Qy 26 PGIVHGVGNFFRAHEAFYVBOILEHAPD---WALVGVGLTSGDRSKKKAEEFKAQDCLY 82
Db 10 PVLLHIGAGSFHRAHQAWYLRVNAVPPGERWILT-VGNIRDDHATLA-ALAAQQGAY 67

Qy 83 SLTEFAPSGKSTVRVMGALRDYLLAPADPEAVLKHVDPAIRIVSMITTEGGYNINETTG 142
Db 68 TLETVTQGERAYETIRSIARVLPWSAALINTGADPACRIYVSTFTVEGGYILDE-HD 126

Qy 143 AFDLENAVAKADLNKPEKPSVFGVYVVEALRRWDAGKATVMSCDNLRHNG---NVAR 199
Db 127 RDVTHPLDAAALRGAR--STLYGALAALLAERRQRGAGPLTLOSCDNLRSNGARFRAGM 184

Qy 200 KAFGLYAKARDEPELAKTTEENATFNGMVDRTPTVSAEIAKKNAAAGLDDDLPLVAED 259
Db 185 RAFL--ALRGDAALLAWFEDANVSCPSAMVDRTPTDDVTRVHAATGVDDRCFVGES 242

Qy 260 FQWVLEDOFADGRPLEKAGVQVGVDTWDEYVKIRMLNAGHVMCLPFGILVGVYVDD 319
Db 243 FTQWIEDNFIAGRPAWEIAGAEIVADVHPVEEAKIRILNATHSCIAWAGTLAGLYIHE 302

Qy 320 AIESELGLNKNLKNKQVITPLKAPSGMTLEGYRDSVISRFSNKAMSQDQTLRTASDCS 379
Db 303 GMRDAIITFRAYDYVDDVIVCL-TPSPDLDERVHDVYLERFNGPNPYVLDLTQNRVAADGFS 361

Qy 380 KVQFWETVTRRAIEDKRLSRIFAGSIYLEMLRG 415
Db 362 KIPGFIAPTLAECFARGADPVATVLPALFLGFLG 397

RESULT 7
DALD_YERPE STANDARD; PRT; 463 AA.
AC P58709;
DT 15-JUN-2002 (Rel. 41, Created)
DT 15-JUN-2002 (Rel. 41, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE D-arabinitol 4-dehydrogenase (EC 1.1.1.11).
GN DALD OR YPO2325.
OS Yersinia pestis.
```

```
OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
OC Yersinia.
OC NCBI_TaxID=632;
RN (1)
RP SEQUENCE FROM N.A.
RC STRAIN-CO-92 / Biovar Orientalis;
RA MEDLINE=21470413; PubMed=11586360;
RA Parkhill J., Wren B.W., Thomson N.R., Titball R.W., Holden M.T.G.,
RA Prentice M.B., Sebaihia M., James K.D., Churcher C., Mungall K.L.,
RA Baker S., Basham D., Bentley S.D., Brooks K., Cerdeno-Tarraga A.M.,
RA Chillingworth T., Cronin A., Davies R.M., Davis P., Dougan G.,
RA Feltwell T., Hamlin N., Holroyd S., Jageis K., Karlyshev A.V.,
RA Leather S., Moule S., Oyston P.C.F., Quail M., Rutherford K.,
RA Simmonds M., Skelton J., Stevens K., Whitehead S., Barrell B.G.;
RT "Genome sequence of Yersinia pestis, the causative agent of plague.";
RL Nature 413:523-527(2001).
CC -!- CATALYTIC ACTIVITY: D-arabinitol + NAD(+) = D-xylulose + NADH.
CC -!- PATHWAY: D-arabinitol catabolism; first step.
CC -!- SIMILARITY: BELONGS TO THE MANNITOL DEHYDROGENASES FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; AJ414152; CAC91130.1;
DR InterPro: IPR000669; Mannitol_dh.
DR Pfam: PF01232; Mannitol_dh; 1.
DR PROSITE: PS00974; MANNITOL_DHGENASE; FALSE_NEG.
KW Oxidoreductase; NAD; Complete proteome.
SQ SEQUENCE 463 AA; 51984 MW; D82E988685B5EDD CRC64;

Query Match 22.68; Score 569; DB 1; Length 463;
Best Local Similarity 36.08; Pred. No. 1.2e-31;
Matches 133; Conservative 71; Mismatches 145; Indels 20; Gaps 10;

Qy 29 VHFVGNGFFRAHEAFYVBOILEHAPD-WALVGVGLTSGDRSKKKAEEFKAQDCLYSLTET 87
Db 13 LHIGAGSFHRAHQAWYLRHLASGDDRTIALANI--RDDAIPLDLTLRAQHGVEVLETV 70

Qy 88 APSGKSTVRVMGALRDYLLAPADPEAVLKHV---DPAIRIVSMITTEGGYNINETTG 143
Db 71 TPGSERQYKETSILRN--IIPWD--KALDHLVAQSKPETRVISFTVTEGGYILD---NQ 123

Qy 144 FDL--ENAAVKAADLNKPEKPSVFGVYVVEALRRWDAGKATVMSCDNLRHNGNVARKA 201
Db 124 FNLQODNSDIOADLRGDCR--TIYGASIRILQORQODKSGPVTLNCDNLRHNGERFRHG 181

Qy 202 FLGYAKARDEPE-LAKWIEENATFNGMVDRTPTVSAEIAKKNAAAGLDDDLPLVAEDF 260
Db 182 LLEFLALRGQSLNWNVTISQTRSPNTWVDRTPTPSADIAQORVLEKTGIVDNAPVWGSEF 241

Qy 261 HOWVLEDOFADGRPLEKAGVQVGVDTWDEYVKIRMLNAGHVMCLPFGILVGVYVDDA 320
Db 242 IQWVIEDDFIAGRPALENVGAEMVTSVLPVEAKIRILNASHSCIAWAGSLIGQCYIHE 301

Qy 321 IEDSELGLNKNLKNKQVITPLKAPSGMTLEGYRDSVISRFSNKAMSQDQTLRTASDCSK 380
Db 302 TOTDAIROMAYDYVDDVIVPSL-TPSPDLDAHYRDVYLERFNSNPYIRDTNORVAADGFSK 360

Qy 381 VQVFWETV 389
Db 361 IPGFITPTL 369

RESULT 8
DALD_KLEPN STANDARD; PRT; 455 AA.
AC O52720;
DT 15-JUN-2002 (Rel. 41, Created)
```


Submitted (SEP-1992) to the EMBL/GenBank/DBJ databases.

[2]
 RL Sequence from N.A.
 RP STRAIN=K12 / MG1655;
 RC MEDLINE=97426617; PubMed=9278503;
 RX Blattnner F.R., Plunkett G. III, Bloch C.A., Perna N.T., Burland V.,
 RA Riley M., Collado-Vides J., Glasner J.D., Rode C.K., Mayhew G.F.,
 RA Gregor J., Davis N.W., Kirkpatrick H.A., Goeden M.A., Rose D.J.,
 RA Mau B., Shao Y.;
 RT "The complete genome sequence of *Escherichia coli* K-12";
 RL Science 277:1453-1474(1997).
 RN [3]
 RP Sequence from N.A.
 RC STRAIN=K12;
 RX MEDLINE=97251357; PubMed=9097039;
 RA Aiba H., Baba T., Fujita K., Hayashi K., Inada T., Isono K., Itoh T.,
 RA Kasai H., Kashimoto K., Kimura S., Kitagawa M., Kitagawa M.,
 RA Makino K., Miki T., Mizobuchi K., Mori H., Mori T., Motomura K.,
 RA Nakade S., Nakamura Y., Nishimoto H., Nishio Y., Oshima T., Saito N.,
 RA Sempel G., Seki Y., Sivasubraman S., Tagami H., Takeda J.,
 RA Takemoto K., Takeuchi Y., Wada C., Yamamoto Y., Horiuchi T.;
 RA "A 570-kb DNA sequence of the *Escherichia coli* K-12 genome
 corresponding to the 28.0-40.1 min region on the linkage map";
 RL DNA Res. 3:363-377(1996).
 RN [4]
 RP Sequence from N.A.
 RC STRAIN=O157:H7 / EDL933 / ATCC 700927;
 RX MEDLINE=21074935; PubMed=11206551;
 RA Perna N.T., Plunkett G. III, Blattnner F.R., Mau B., Glasner J.D.,
 RA Rose D.J., Mayhew G.F., Evans P.S., Gregor J., Kirkpatrick H.A.,
 RA Posfai J., Hackett J., Klink S., Boutin A., Shao Y., Miller L.,
 RA Grotbeck E.J., Davis N.W., Lim A., Dimalanta E.T., Potamouls K.,
 RA Apodaca J., Anantharaman T.S., Lin J., Yen G., Schwartz D.C.,
 RA Welch R.A., Blattnner F.R.;
 RT "Genome sequence of enterohaemorrhagic *Escherichia coli* O157:H7";
 RL Nature 409:529-533(2001).
 RN [5]
 RP Sequence from N.A.
 RC STRAIN=O157:H7 / RIMD 0509952;
 RX MEDLINE=21156231; PubMed=11258796;
 RA Hayashi T., Makino K., Ohnishi M., Kurokawa K., Ishii K., Yokoyama K.,
 RA Han C.-G., Ohtsubo E., Nakayama K., Murata T., Tanaka M., Tobe T.,
 RA Iida T., Takami H., Honda T., Sasakawa C., Ogasawara N., Yasunaga T.,
 RA Kihara S., Shiba T., Hattori M., Shinagawa H.;
 RT "Complete genome sequence of enterohaemorrhagic *Escherichia coli* O157:H7 and genomic comparison with a laboratory strain K-12";
 RL DNA Res. 8:11-22(2001).
 RN [6]
 RP Sequence of 1-45 from N.A., and sequence of 1-4.
 RC STRAIN=K12;
 RX MEDLINE=86280350; PubMed=3525742;
 RA Blanco C., Mata-Gilsinger M.;
 RT "A DNA sequence containing the control sites for the *uxaB* gene of
Escherichia coli";
 RL J. Gen. Microbiol. 132:697-705(1986).
 CC -!- CATALYTIC ACTIVITY: D-altronate + NAD(+) = D-tagaturonate + NADH.
 CC -!- PATHWAY: Glucuronate pathway.
 CC -!- INDUCTION: BY GALACTURONATE, TAGATURONATE AND FRUCTURONATE. ITS
 CC EXPRESSION IS SUBJECTED TO CATABOLITE REPRESSION BY GLUCOSE.
 CC -!- SIMILARITY: SOME, TO THE MANNITOL DEHYDROGENASE FAMILY.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 CC EMBL; D13327; BAA02586.1; -
 CC EMBL; AE000249; AAC74594.1; -
 CC EMBL; D90795; BAA15204.1; -
 CC EMBL; AE005352; AAG56245.1; -

DR EMBL; AP002557; BAB35551.1; -
 DR EMBL; M15737; AAA24757.1; -
 DR PIR; A46549; A46549.
 DR EcoGene; EG11065; *uxaB*.
 DR InterPro; IPR000669; Mannitol_dh.
 DR Pfam; PF01232; Mannitol_dh; 1.
 KW Oxidoreductase; NAD; Complete proteome.
 FT CONFLICT 332 332. P -> L (IN REF. 2).
 SQ SEQUENCE 483 AA; 54808 MW; AFA4677CF35BC2E0 CRC64;
 Query Match 11.38; Score 285; DB 1; Length 483;
 Best Local Similarity 23.68; Pred. No. 3.le-12;
 Matches 120; Conservative 89; Mismatches 200; Indels 100; Gaps 21;
 Qy 28 IVHFGVGNFFRAHFAFYVEQILEHAP-DWAIVGVGLTGSRSKKKAE-----EFAKQDCL 81
 Db 18 IIOFGEGNFLRAFVDMQIDLLNHTDLSGVVV-----RPIETSFPPSLSTQDGL 68
 Qy 82 YS-----LTETAPSGKSTVRVMGALROYLLAPAEAVLKLHVDPAIRIVSWTITEGGYN 136
 Db 69 YTTIRGLNEKG-EAVSDARLIRSVMREISVSEYDFELKLAHNPMPFRVFSNTTEAGIS 127
 Qy 137 INEYTGAFDLENAVKADLKNPEKPSVTFVGVVEALRRRWDAAGKAFVMSCDNLRHNGN 196
 Db 128 YH-AGDKFD-DAPAVSYPAKUTRLLEFERSHFGALDKGN-----IIPCELDYNGD 178
 Qy 197 VARKAFLGYAK--ARDPELAKWIEENATFPMGMVDRIPTVSAETAKKLNAAAGLDDDL 254
 Db 179 ALRELVLRYAQEWALPEAFIOMLQOANSFCSTLVDRIVTGYPRDEVAKLEELGYHDGFL 238
 Qy 255 LVAEDFHQVLE-----DOFADGRPLEKAGVQMGVDVTDVYVYKIRMLNAGHV 303
 Db 239 DTAEHFYLFTVQGPKSALTELRDKY-----PL---NVLLVDDIKPKYKRVAILNGAHT 290
 Qy 304 MLCPPGILVGVENVDDAIEDSELNKNLYLNKQVITPLKAPSGMTLECYRDSVSIRSN 363
 Db 291 ALVPVAFQAGLDVGEAMNDAEICAFVEKAYEIIIPVLDLPRD-ELESFASAVTGRFN 349
 Qy 364 KAMSDQTLRIASDGCCKVQVFWTETVRRRAIEDKRD-----LSRIAFGIASYLEMLRGREK 419
 Db 350 PYIKHQLLSIALNGMTK---FTRILPOLLGAKANGTLPARTLFAALAATFYRG-ERN 405
 Qy 420 GGTYESSEPTVGDREW-----KLAKADDFESSL-KLPAPFG 454
 Db 406 GETY----PVODDAHLERYQQLWSQHRDRVIGTQELVAIVLAEKDHWEDLTQVP---G 458
 Qy 455 WRDLDTSELQOKVIVLRKIIIEKGVKAAI 483
 Db 459 LVEQVANDLD-----AILEKGMREAV 479
 RESULT 11
 POR_GLUJUX
 ID POR_GLUJUX STANDARD: PRT: 50 AA.
 AC P80354;
 DT 01-NOV-1995 (Rel. 32, Created)
 DT 01-NOV-1995 (Rel. 32, Last sequence update)
 DT 16-OCT-2001 (Rel. 40, Last annotation update)
 DE Polyl:NADP oxidoreductase (EC 1.1.1.-) (Fragment).
 GN POR.
 OS Gluconobacter oxydans (Gluconobacter suboxydans).
 OC Bacteria; Proteobacteria; alpha subdivision; Acetobacteraceae;
 OC Gluconobacter.
 OX NCBI_TaxID=442;
 RN [1]
 RP SEQUENCE.
 RC STRAIN=DSM 350;
 RA Klaseen R.;
 RL Thesis (1994), Heinrich-Heine University / Duesseldorf, Germany.
 CC -!- SUBCELLULAR LOCATION: Cytoplasmic.
 DR InterPro: IPR000669; Mannitol_dh.
 DR Pfam; PF01232; Mannitol_dh; 1.
 KW Oxidoreductase; NADP.

Db	4	LHFCAGNIGRGFTG	YLLYSKSYETTF	VDKVVDDINKYKRYTVITLSTSKNKEV	63
Qy	76	KAQDCLYSLTETAP	SGKSTVRVMGAL	RDYLLAPADPEAVLKLHVDPAIRIVSVTTEGGY	135
Db	64	RAVNL	-----KDSVALEKEVLE	-----ADLIITSLGL	90
Qy	136	NINETTCAGFDLE	NAVKADLKNPEK	PSVFGYVVEALRRRWDAAGKRAFTVMSCDNLRHNG	195
Db	91	NNLKSTG	-----ELLRGF	-----LKKRSEINDRPDLIIACENALFAS	127
Qy	196	NVARKAFGLYAKA	RDELAKWIENAT	FPNGMVDRIPTPTVSABIAKKLNAASGLDDDLPL	255
Db	128	DVLKAILDGA	-----DEELKYLEKSVG	FPNCTVDRIVPNV-----DIEKEL	174
Qy	256	VAEDFHOWVL	EDQFADGRPPLE	KAGVQMVGDDTWEEV-----KTRMLNAGHVMLC	306
Db	175	AVEDFYEW	-----DIEKNKKKIN	KIIGAIEYVEKLDPLYLERKLFLLNGAHATIA	223
Qy	307	FPGLTVGVENV	DDAIEDSELLGN	LKNLNDKVI-----PTLKAPSG-----MTLEGYRDS	356
Db	224	YLGVLKGYKI	HEAIKDK	-----INKIIVGFHSEAVQALSEKHKIDIOILKEYSNK	275
Qy	357	VISRFSNKAMS	DDQTLRIASDC	GCKVQV--FWTETVRRRAIEDKDLRSRIAFGSIASYLEMLR	414
Db	276	LLKRFENEYLK	DDVSRVGRDPM	RKRLLSSNDRLITPLKCCDLKIDFTNILFGVAS-----	329
Qy	415	GRDEKGGTYE	SESSEPTYGDAE	WKLAKADDFESSLKLPAFDGWRDLDTSELDDOKVIVLRKII	474
Db	330	-----GYLVNKE	-----	-----DEKAQGIQNI	348
Qy	475	REKGVKAAI	483		
Db	349	TKEGIKKAI	357		
RESULT 13					
MTLD	ECOLI	STANDARD;	PRT;	382 AA.	
ID	MTLD_ECOLI				
AC	P09424;				
DT	01-MAR-1989 (Rel. 10, Created)				
DT	01-OCT-1994 (Rel. 30, Last sequence update)				
DT	16-OCT-2001 (Rel. 40, Last annotation update)				
DE	Mannitol-1-phosphate 5-dehydrogenase (EC 1.1.1.17).				
GN	MTLD OR B3600.				
OS	Escherichia coli.				
OC	Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;				
OC	Escherichia.				
OX	NCBI_TaxID=562;				
RN	[1]				
RP	SEQUENCE FROM N.A.				
RC	STRAIN=K12;				
RX	MEDLINE=88288055; PubMed=3135464;				
RA	Davis T., Yamada M., Elgort M., Sailer M.H. Jr.;				
RT	"Nucleotide sequence of the mannitol (mtl) operon in Escherichia				
RT	coli.";				
RL	Mol. Microbiol. 2:405-412(1988).				
RL	[2]				
RP	SEQUENCE FROM N.A., AND REVISIONS.				
RC	STRAIN=K12;				
RX	MEDLINE=91186837; PubMed=1964486;				
RA	Jiang W., Wu L.F., Tomich J., Sailer M.H. Jr., Nicausa W.G.;				
RT	"Corrected sequence of the mannitol (mtl) operon in Escherichia				
RT	coli.";				
RL	Mol. Microbiol. 4:2003-2006(1990).				
RL	[3]				
RP	SEQUENCE FROM N.A.				
RC	STRAIN=K12 / MG1655;				
RX	MEDLINE=94316500; PubMed=8041620;				
RA	Sofia H.J., Burland V., Daniels D.L., Plunkett G. III, Blattner F.R.;				
RT	"Analysis of the Escherichia coli genome. V. DNA sequence of the				
RT	region from 76.0 to 81.5 minutes.";				
RL	Nucleic Acids Res. 22:2576-2586(1994).				
RL	[4]				

GenCore version 5.1.4_p5_4578
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: May 8, 2003, 16:28:22 ; Search time 38 Seconds
(without alignments)
2629.812 Million cell updates/sec

Title: US-09-926-163B-2

Perfect score: 2516

Sequence: 1 MITRETKSLPANVQAPPYD.....KVIVLRKIIRKGVKAAIPA 485

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 671580 seqs, 206047115 residues

Total number of hits satisfying chosen parameters: 671580

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

SPTREMBL_21.*

- 1: sp_archaea.*
- 2: sp_bacteria.*
- 3: sp_fungi.*
- 4: sp_human.*
- 5: sp_invertebrate.*
- 6: sp_mammal.*
- 7: sp_mhc.*
- 8: sp_organelle.*
- 9: sp_phage.*
- 10: sp_plant.*
- 11: sp_rodent.*
- 12: sp_virus.*
- 13: sp_vertibrate.*
- 14: sp_unclassified.*
- 15: sp_rvirus.*
- 16: sp_bacteriap.*
- 17: sp_archaeap.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Match %	Length	DB	ID	Description
1	2516	100.0	485	2	Q9KWR5	Q9kwr5 gluconobact
2	2143	85.2	485	2	Q8VUU5	Q8vuu5 gluconobact
3	922	36.6	491	16	Q911D6	Q91ld6 pseudomonas
4	888	35.3	493	2	O88355	O88355 pseudomonas
5	797	31.7	492	16	Q98D06	Q98d06 rhizobium l
6	779	31.0	487	16	Q92TQ9	Q92tg9 rhizobium m
7	775.5	30.8	486	16	O8X516	O8x516 escherichia
8	769.5	30.6	486	16	O8X506	O8x506 escherichia
9	768	30.5	478	2	O68111	O68111 rhodobacter
10	760	30.2	494	16	Q92N07	Q92n07 rhizobium m
11	759.5	30.2	488	16	O8Z626	O8z626 salmonella
12	755.5	30.0	488	16	O8ZP18	O8zpi8 salmonella
13	753.5	29.9	490	16	O8ZGK9	O8zgk9 yersinia pe
14	746.5	29.7	488	16	O8XE83	O8xe83 escherichia
15	730	29.0	490	16	O8XFB5	O8xfb5 salmonella
16	730	29.0	525	16	Q8U7J8	Q8u7j8 agrobacteri

17	721	28.7	458	16	Q9A877	Q9a877 caulobacter
18	700.5	27.8	502	16	Q8YCO0	Q8ycq0 brucella me
19	678.5	27.0	499	16	Q8EWM0	Q8ew0 rhizobium l
20	650.5	25.9	490	16	Q8UA45	Q8ua45 agrobacteri
21	587.5	23.4	490	16	Q8ZM58	Q8zm58 salmonella
22	556.5	22.1	497	16	Q98JL8	Q98jl8 rhizobium l
23	518.5	20.6	439	2	Q9XBD5	Q9xbd5 amycolatops
24	405	16.1	539	16	Q8R8Q5	Q8r8q5 thermoanaer
25	392.5	15.6	285	2	Q93KC4	Q93kc4 erwinia chr
26	371	14.7	539	16	Q9WXS3	Q9wxs3 thermotoga
27	365.5	14.5	544	16	Q9CF49	Q9cf49 lactococcus
28	343.5	13.7	482	16	Q97167	Q97167 clostridium
29	331	13.2	480	16	Q34354	Q34354 bacillus su
30	299.5	11.9	512	16	Q9KFI7	Q9kfi7 bacillus ha
31	288	11.4	483	16	Q8ZIC5	Q8zic5 yersinia pe
32	238	9.5	386	16	Q8RCS0	Q8rcs0 thermoanaer
33	214	8.5	366	2	Q9KH89	Q9kh89 escherichia
34	213	8.5	366	2	Q9KH91	Q9kh91 escherichia
35	209	8.3	366	2	Q9KH92	Q9kh92 escherichia
36	208	8.3	366	2	Q9K2M3	Q9k2m3 escherichia
37	207	8.2	366	2	Q9K2H8	Q9k2h8 escherichia
38	206	8.2	366	2	Q9KH90	Q9kh90 escherichia
39	206	8.2	382	16	Q8XDG9	Q8xdg9 escherichia
40	205	8.1	366	2	Q9K2Y2	Q9k2y2 escherichia
41	205	8.1	366	2	Q9K2U2	Q9k2u2 escherichia
42	202.5	8.0	387	16	Q8Z9X0	Q8z9x0 yersinia pe
43	202	8.0	366	2	Q9K2V6	Q9k2v6 escherichia
44	202	8.0	366	2	Q9KH93	Q9kh93 escherichia
45	201	8.0	366	2	Q9KH88	Q9kh88 escherichia

ALIGNMENTS

RESULT 1

Q9KWR5 ID Q9KWR5 PRELIMINARY; PRT; 485 AA.
AC Q9KWR5;
DT 01-OCT-2000 (TREMBLrel. 15, Created)
DT 01-OCT-2000 (TREMBLrel. 15, Last sequence update)
DT 01-OCT-2001 (TREMBLrel. 18, Last annotation update)
DE Sorbitol dehydrogenase.
OS Gluconobacter oxydans (Gluconobacter suboxydans).
OC Bacteria; Proteobacteria; alpha subdivision; Acetobacteraceae;
OC Gluconobacter.
OX NCBI_taxID=442;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=G624;
RA Shibata T., Ichikawa C., Matsuura M., Takata Y., Noguchi Y., Saito Y.,
RA Yamashita M.;
RT "Cloning of a gene for D-Sorbitol dehydrogenase from Gluconobacter
oxydans G624 and expression of the gene in Pseudomonas putida
IFO3738".
RL J. Biosci. Bioeng. 89:463-468(2000).
DR EMBL; AB028937; BAA99414.1;
DR InterPro; IPR000669; Mannitol_dh.
DR Pfam; PF01232; Mannitol_dh; 1.
DR PRINTS; PR00084; MTLDHDHGNASE.
SQ SEQUENCE 485 AA; 53642 MW; A7C363D728EEA7DC CRC64;

Query Match 100.0%; Score 2516; DB 2; Length 485;
Best Local Similarity 100.0%; Pred. No. 1.5e-173;
Matches 485; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 MITRETKSLPANVQAPPYDIDGKPGIVHFGVGNFFRAHEAFYVEQILEHAPDWAIVGV 60

Db 1 MITRETKSLPANVQAPPYDIDGKPGIVHFGVGNFFRAHEAFYVEQILEHAPDWAIVGV 60

Qy 61 GLTGSDRSKKAEERKADCLYSLTETAPSGKSTVRVVGALRDYLLAPADPAVLKHLVD 120

Db 61 GLTGSDRSKKAEERKADCLYSLTETAPSGKSTVRVVGALRDYLLAPADPAVLKHLVD 120

```
QY 121 PAIRIVSMITTEGGYNINETTGAFDLENAAVKADLNKPEKSTVFYGVVEALRRWDAGG 180
|||||
Db 121 PAIRIVSMITTEGGYNINETTGAFDLENAAVKADLNKPEKSTVFYGVVEALRRWDAGG 180

QY 181 KFTVMSCDNLRHNGNVARKAFLGYAKARDPELAKWIEENATFPNGWDRITPTVSAAETA 240
|||||
Db 181 KFTVMSCDNLRHNGNVARKAFLGYAKARDPELAKWIEENATFPNGWDRITPTVSAAETA 240

QY 241 KKLNAASGLDDDLPLVAEDFHQWLEDFADGRPPLEKAGVOMGVGDVTDWEYVKIRMLNA 300
|||||
Db 241 KKLNAASGLDDDLPLVAEDFHQWLEDFADGRPPLEKAGVOMGVGDVTDWEYVKIRMLNA 300

QY 301 GHVMLCFPGILVGYENVDDAIEDSELGLNKNYLNKDVITPLKAPSGMTLEGYRDSVISR 360
|||||
Db 301 GHVMLCFPGILVGYENVDDAIEDSELGLNKNYLNKDVITPLKAPSGMTLEGYRDSVISR 360

QY 361 FSNKAMSDOTLRIASDGCSSKQVFWTETVRRRAIEDKRLSRIAFGSIASYLEMLRGRDEKG 420
|||||
Db 361 FSNKAMSDOTLRIASDGCSSKQVFWTETVRRRAIEDKRLSRIAFGSIASYLEMLRGRDEKG 420

QY 421 GTYESSEPTYGDAEWKLAKADDFESSLKLPAFDGWRDLDTSELDOKVIVLRKIIREKGVK 480
|||||
Db 421 GTYESSEPTYGDAEWKLAKADDFESSLKLPAFDGWRDLDTSELDOKVIVLRKIIREKGVK 480

QY 481 AAIPA 485
|||||
Db 481 AAIPA 485
```

RESULT 2

```
Q8VUU5 ID Q8VUU5 PRELIMINARY; PRT; 485 AA.
AC Q8VUU5;
DT 01-MAR-2001 (TREMBlrel. 20, Created)
DT 01-MAR-2002 (TREMBlrel. 20, Last sequence update)
DT 01-JUN-2002 (TREMBlrel. 21, Last annotation update)
DE L-sorbose reductase.
SR.
OS Gluconobacter oxydans (Gluconobacter suboxydans).
OC Bacteria; Proteobacteria; alpha subdivision; Acetobacteraceae;
OC Gluconobacter.
OX NCBI_TaxID=442;
RN [1]
RP SEQUENCE FROM N.A.
RX STRAIN=IFO3291;
RX MEDLINE=21650687; PubMed=11790761;
RA Shiojoh M., Tazoe M., Hoshino T.;
RT "NADPH-Dependent L-Sorbose Reductase Is Responsible for L-Sorbose
RT Assimilation in Gluconobacter suboxydans IFO 3291.";
RL J. Bacteriol. 184:861-863(2002).
DR EMBL; AB063188; BAB83933.1;
DR InterPro: IPR000669; Mannitol_dh.
DR Pfam: PF01232; Mannitol_dh; 1.
DR PRINTS; PR00084; MTLDHDRGNASE.
SQ SEQUENCE 485 AA; 53541 MW; 883EC4956B3C6CF0 CRC64;
```

```
Query Match 85.2%; Score 2143; DB 2; Length 485;
Best Local Similarity 84.5%; Pred. No. 1.3e-146;
Matches 410; Conservative 30; Mismatches 45; Indels 0; Gaps 0;
```

```
QY 1 MITRETLKSLPANVQAPPYDIDGKPGIVHFGVGNFFRAHEAFYVEQILEHAPDAIVGV 60
|||||
Db 1 MITRETLKSLPANVQAPPYDIDGKPGIVHFGVGNFFRAHEAFYVEQILEHAPDAIVGV 60

QY 61 GLTGSDRSKKKAEKFAKQDCLSLTETAPSGKSTVRVMGALRDYLLAPADPEAVLKLVD 120
|||||
Db 61 GLTGSDRSKKKAEKFAKQDCLSLTETAPSGKSTVRVMGALRDYLLAPADPEAVLKLVD 120

QY 121 PAIRIVSMITTEGGYNINETTGAFDLENAAVKADLNKPEKSTVFYGVVEALRRWDAGG 180
|||||
Db 121 PAIRIVSMITTEGGYNINETTGAFDLENAAVKADLNKPEKSTVFYGVVEALRRWDAGG 180

QY 181 KFTVMSCDNLRHNGNVARKAFLGYAKARDPELAKWIEENATFPNGWDRITPTVSAAETA 240
|||||
```

```
Db 181 KFTVMSCDNLRHNGNVARKAFLGYAKARDPELAKWIEENATFPNGWDRITPTVSADIA 240
|||||
QY 241 KKLNAASGLDDDLPLVAEDFHQWLEDFADGRPPLEKAGVOMGVGDVTDWEYVKIRMLNA 300
|||||
Db 241 KKLNEASGLDDDLPLVAEDFHQWLEDFADGRPALEKAGVQFVGVDVTDYEHVKIRMLNA 300
|||||
QY 301 GHVMLCFPGILVGYENVDDAIEDSELGLNKNYLNKDVITPLKAPSGMTLEGYRDSVISR 360
|||||
Db 301 GHVMLCFPGILVGYENVDDAIEDSELGLNKNYLNKDVITPLKAPSGMTLEGYRDSVISR 360

QY 361 FSNKAMSDOTLRIASDGCSSKQVFWTETVRRRAIEDKRLSRIAFGSIASYLEMLRGRDEKG 420
|||||
Db 361 FSNPAMADOTLRIASDGCSSKQVFWTETVRRRAIEKRLSRIAFGSIASYLEMLRGRDEKG 420

QY 421 GTYESSEPTYGDAEWKLAKADDFESSLKLPAFDGWRDLDTSELDOKVIVLRKIIREKGVK 480
|||||
Db 421 GTYEFEPFTFGDNHKTAKADDFESSLKLPAFDGWRDLDTSELDOKVIVLRKIIREKGVK 480

QY 481 AAIPA 485
|||||
Db 481 AAIPA 485

PRELIMINARY; PRT; 491 AA.
Q91LD6 ID Q91LD6 PRELIMINARY; PRT; 491 AA.
AC Q91LD6;
DT 01-MAR-2001 (TREMBlrel. 16, Created)
DT 01-MAR-2002 (TREMBlrel. 16, Last sequence update)
DT 01-MAR-2002 (TREMBlrel. 20, Last annotation update)
DE Mannitol dehydrogenase.
GN MTLR OR PA2342.
OS Pseudomonas aeruginosa.
OC Bacteria; Proteobacteria; gamma subdivision; Pseudomonadaceae;
OC Pseudomonas.
OX NCBI_TaxID=287;
RN [1]
RP SEQUENCE FROM N.A.
RX STRAIN=ATCC 15692 / PAOI;
RX MEDLINE=20437337; PubMed=10984043;
RA Stover C.K., Pham X.-Q.T., Erwin A.L., Mizoguchi S.D., Warren P.,
RA Hickey M.J., Brinkman F.S.L., Hufnagle W.O., Kowalik D.J., Lagrou M.,
RA Garber R.L., Goulet L., Tolentino E., Westbrock-Wadman S., Yuan Y.,
RA Brody L.A., Coulter S.N., Folger K.R., Kas A., Larbig K., Lim R.M.,
RA Smith K.A., Spencer D.H., Wong G.K.-S., Wu Z., Paulsen I.T.,
RA Reizer J., Saier M.H., Hancock R.E.W., Lory S., Olson M.V.;
RT "Complete genome sequence of Pseudomonas aeruginosa PAOI, an
RT opportunistic pathogen.";
RL Nature 406:959-964(2000).
DR EMBL; AE004660; AAG05730.1;
DR InterPro: IPR000669; Mannitol_dh.
DR Pfam: PF01232; Mannitol_dh; 1.
DR PRINTS; PR00084; MTLDHDRGNASE.
DR PROSITE; PS00974; MANNITOL_DHGENSE; 1.
KW Complete proteome.
SQ SEQUENCE 491 AA; 54350 MW; 8729A9F7E9841C39 CRC64;
```

```
Query Match 36.6%; Score 922; DB 16; Length 491;
Best Local Similarity 45.2%; Pred. No. 2.6e-58;
Matches 199; Conservative 66; Mismatches 171; Indels 4; Gaps 2;
```

```
QY 2 ITRETLKSLPANVQAPPYDIDGKPGIVHFGVGNFFRAHEAFYVEQILEHAPDAIVGV 59
|||||
Db 3 LNROHLPLLATAVAPSYDPAQIRGIVHGVGFHRAHQAYTDLNLRGALDWAICG 62
|||||
QY 60 VGLTGSDRSKKKAEKFAKQDCLSLTETAPSGKSTVRVMGALRDYLLAPADPEAVLKLVD 119
|||||
Db 63 AGLRSDRAMHDA--LAAQDYLYTLYELCDQPDTEVRVIGATSGMLLAEDGAELKLA 120
|||||
QY 120 PAIRIVSMITTEGGYNINETTGAFDLENAAVKADLNKPEKSTVFYGVVEALRRWDAG 179
|||||
Db 121 EPAIRIVSLTITEGGYCIDDGSGEFLAELPLVRHDLANPRTPRGVGFCEALRRRDTG 180
|||||
```



```

Query Match      30.8%;   Score 775.5;   DB 16;   Length 486;
Best Local Similarity 36.4%;   Pred. No. 9.9e-48;
Matches 177;   Conservative 87;   Mismatches 205;   Indels 17;   Gaps 9;

Qy 9  SLPANVOAPPYDIDIGIKPIGVHFGVGNFFRAHEAFYVEQIILEA-PDKWAIVGVGL-TGSD 66
    : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 8  NLP--VARPSWDHSRLSRIVHLGGCAFRAHQALYTHHLLSTDSDGWGICEVNLMPGND 65

Qy 67 RSKKAAEFKQADCLYSLTETAPSGKSTVRVNGALRDYLLAPADP-EAVLKHLDVDPATRI 125
Db 66 R--VLIIENLKKQQLLYTVAERG-AESTELKIIGSMKEALHPEDIDCEGILNAMAQPQTAI 122

Qy 126 VSMITTEGGYNIETGAFDLENAVAKDLKNPEKPSVFGVYVEALRRRDADCGKAFTV 185
Db 123 VSLIVTEKGYCADAAASGQDLNPLIKHDLNPTAPKSAIGTIYVEALRLREKKGUKAFTV 182

Qy 186 MSCDLNRHNGVAKAFGLGYAKADPELAKWIEENATPNGMVDRITPTVSAEIAKKLNA 245
Db 183 MSCDNRENGHVAKVAVLGLAQARDPQLAAWIEENVTPCTMVDRIVPAATPETLQELAD 242

Qy 246 ASGLDDDLPLVAEDFQHWLEQFADGPPLEKAGVQMVGDVTWDEYVKIRMLNAGHVML 305
Db 243 QLVGYDPCATACEFRQMWIEDNFVNGPDKDKVGAQFVADVFPFEMKMLRMLNGSHSL 302

Qy 306 CFPGLVGYENVDDAIEDSELLGNLKNLYNKDVIPTLKAPSGMTLEGYRDSVIRSFNSKA 365
Db 303 AYLGLVGYETIADTMNPDYRKAALFMWQEQAPTLSMPEGTDLNAYATLLIERFSNPS 362

Qy 366 MSDOTLRASDGSCKVQVFTWETVRRALDLSRIAGFASYLEMLRGRDEKGGYVES 425
Db 363 LRHRTWQIAMDGSQKLPQRLDPVRLHLQNGSGSWRHLALGVAAGMVRVYTGQVDEQGNADIV 422

Qy 426 SEPTYGAENWLKADDFESSLLKAPFDGWRDLDTSELDQK-----VIVLRKTIIREKGV 479
Db 423 VDPML--AEFOKINA-QYOGADRVKALLGLSGIFADDPQNAFVGAVTAAYQQLCERGA 479

Qy 480 KAAIPA 485
Db 480 RECVA 485

```

```

RESULT 8
Q8X506
ID Q8X506 PRELIMINARY; PRT; 486 AA.
AC Q8X506;
DT 01-MAR-2002 (TrEMBLrel. 20, Created)
DT 01-MAR-2002 (TrEMBLrel. 20, Last sequence update)
DT 01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
DE Putative oxidoreductase.
GN YDFI OR Z2155 OR ECS2151.
OS Escherichia coli O157:H7.
OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
OC Escherichia.
OX NCBI_TaxID=83334;
RN [1]
RP SEQUENCE FROM N.A.
RX STRAIN=O157:H7 / EDL933 / ATCC 700927;
RX MEDLINE=21074935; PubMed=11206551;
RA Perse N.T., Plunkett G. III, Burland V., Mau B., Glasner J.D.,
RA Rose D.J., Mayhew G.F., Evans P.S., Gregor J., Kirkpatrick H.A.,
RA Posfal G., Hackett J., Klink S., Boutin A., Shao Y., Miller L.,
RA Grothbeck E.J., Davis N.W., Lim A., Dimalanta E.T., Potamocousis K.,
RA Apodaca J., Anantharaman T.S., Lin J., Yen G., Schwartz D.C.,
RA Welch R.A., Blattner F.R.;
RT "Genome sequence of enterohaemorrhagic Escherichia coli O157:H7." ;
RL Nature 409:529-533(2001).
RN [2]
RP SEQUENCE FROM N.A.
RX STRAIN=O157:H7 / RIMD 0509952;
RX MEDLINE=21156231; PubMed=11258796;
RA Hayashi T., Makino K., Ohnishi M., Kurokawa K., Ishii K., Yokoyama K.,
RA Han C.-G., Ohtsubo E., Nakayama K., Murata T., Tanaka M., Tobe T.,

```

[illegible]

```

RESULT 9
O68111
ID O68111 PRELIMINARY; PRT; 478 AA.
AC O68111;
DT 01-AUG-1998 (TREMBLrel. 07, Created)
DT 01-AUG-1998 (TREMBLrel. 07, Last sequence update)
DT 01-JUN-2001 (TREMBLrel. 17, Last annotation update)
DE Mannitol 2-dehydrogenase (EC 1.1.1.67).
OS Rhodobacter capsulatus (Rhodopseudomonas capsulata).
OC Bacteria; Proteobacteria; alpha subdivision; Rhodobacter group;
OC Rhodobacter.
OX NCBI_TaxID=i061;
RX [1]
RN SEQUENCE FROM N.A.
RP STRAIN=SB1003;
RC MEDLINE=97404404; PubMed=9256491;
RA Vitek C., Paces V., Maltsev N., Paces J., Haselkorn R., Fonstein M.;
RT "Sequence of a 189-kb segment of the chromosome of Rhodobacter

```

RT	capsulatus Sbl003.;	RT	"Analysis of the chromosome sequence of the legume symbiont
RL	Proc. Natl. Acad. Sci. U.S.A. 94:9384-9388(1997).	RL	Sinorhizobium meliloti strain 1021.;
DR	EMBL; AF010496; AAC16201.1;	DR	Proc. Natl. Acad. Sci. U.S.A. 98:9877-9882(2001).
DR	InterPro: IPR000669; Mannitol_dh.	DR	EMBL; AL591790; CAC47019.1;
DR	Pfam: PF01232; Mannitol_dh; 1.	DR	InterPro: IPR000669; Mannitol_dh.
DR	PRINTS: PR00084; MTLDRHGNASE.	DR	Pfam: PF01232; Mannitol_dh; 1.
DR	PROSITE; PS00974; MANNITOL_DHGENASE; 1.	DR	PROSITE; PS00974; MANNITOL_DHGENASE; UNKNOWN_1.
KW	Oxidoreductase.	KW	Oxidoreductase; Complete proteome.
SQ	SEQUENCE 478 AA; 51442 MW; 8427F89B6F2A4F38 CRC64;	SQ	SEQUENCE 494 AA; 53822 MW; 7533287D8EF58F24 CRC64;
<p>Query Match 30.5%; Score 768; DB 2; Length 478; Best Local Similarity 36.4%; Pred. No. 3.3e-47; Matches 176; Conservative 84; Mismatches 202; Indels 22; Gaps 9;</p>		<p>Query Match 30.28; Score 760; DB 16; Length 494; Best Local Similarity 36.18; Pred. No. 1.3e-46; Matches 176; Conservative 83; Mismatches 208; Indels 20; Gaps 8;</p>	
Qy	10 LPANQAAPPYDIDGKPGIVHFGVGNFFRAHEAFVEQI--LEHAPDAIIVGVGTGSDR 67	Qy	6 TLKSLPANQAAPPYDIDGKPGIVHFGVGNFFRAHEAFVEQI--LEHAPDAIIVGVGTG 63
Db	1 MPVPAAPRYDRAQTPTGIVHIGLGNFIRAHQAVYLDLFLGLGEGHDWALIGAGVRAFDA 60	Db	9 TLDVAVKATAGVNPGRHDLRAGIVHFGVGNFIRAHQAVYLDLFLGLGRDRWAIIGAGVL 68
Qy	68 SKKKAEEFKADOCYLSLTETAPSGKSTVVMGALRDYLLAPADPAEVLKHLVDPAIRVS 127	Qy	64 GDSRKKKAEEFKADOCYLSLTETAPSGKSTVVMGALRDYLLAPADPAEVLKHLVDPAI 123
Db	61 AMR--EVLLAODGLSSVTELAP-GAHSARVICAMIDEVAVQFGNAALATWARPPIRVS 117	Db	69 PSD--KVMRDKLEAQDFLTTVVE-QDNNRTGAHVTGAMIAV-LEPGDTPAIVQAQLASPLI 124
Qy	128 MTITEGGYNINETTCAFDLENAAVKADLNKPEKPTVFGYVVEALRRWDAGGKFTVMS 187	Qy	124 RIVSMITTEGGYNINETTCAFDLENAAVKADLNKPEKPTVFGYVVEALRRWDAGGKRAF 183
Db	118 LTVTEGGYIIDNTGTFSPDHPDIOQAARPDAPDTAFGAIIAALKMRAAGVAPFTVMC 177	Db	125 RIVSLTITEGGYIIDPASGVDFDPAHVAIVEDARDPAKPTVFGLLAGLAERRAKGIPPF 184
Qy	188 CNLRHNGNVARKAFGLYAKARDPELAKWIENATFPNGMVDRIPTTVSAETAKKLNAAS 247	Qy	184 TWMSCDNLRHNGNVARKAFGLYAKARDPELAKWIENATFPNGMVDRIPTTVSAETAKKL 243
Db	178 CNDVPHNGHVTDAVVGRLARLEDAELADWIHASVAFPSMWDRITP-ATGDERAMAADL 236	Db	185 TIMSCDNLPGNGEVTHAAVSGRLASDPGFADWDIANDVAFNGMVDRIPTPATGAREIGIV 244
Qy	248 GLDDDLPLVAEDFHQWLEDOFADGRPPLEKAGVQMGVDVTDWEYVKIRMLNAGHVMICF 307	Qy	244 NAASGLDDDLPLVAEDFHQWLEDOFADGRPPLEKAGVQMGVDVTDWEYVKIRMLNAGHV 303
Db	237 GLDDAAPVTCETFRQWTEIDFKPAGRPALKEKVGVTFTPHVDKFTMKIRILNGHAIAY 296	Db	245 ASQYIGDDAWPVCFEEFKQWLEDRFPQGRPALEEVGVQFVDPVAPYEHMKIRIILNGGHA 304
Qy	308 PGILVYENVDDAIEDSELLGNLKNYLNKQVITPLKAPSGMTLEGYRDSVIRFSNKAMS 367	Qy	304 MLCPPGILVYENVDDAIEDSELLGNLKNYLNKQVITPLKAPSGMTLEGYRDSVIRFSN 363
Db	297 PGGLADIATFVHEAMADPOIRAFNLVLAETVLPVPPVPGDLNDYKALITIERFSNPEVA 356	Db	305 AIATPAALLDHFVHEAMEEPLIFRAFLSKLEHDELIIPVPPVDDTLKDYKLIETFRSN 364
Qy	368 DOTLRASDGSKVQVFWTETVRRRAIEDKRLSRIAFAFGIASYLEMLGRDEKGGYSESS 427	Qy	364 KAMSDOTLRASDGSKVQVFWTETVRRRAIEDKRLSRIAFAFGIASYLEMLGRDEKGGTY 423
Db	357 DIVRRLCLDGSNRQPKFIIPSIIRDRIATGSGFDGLVLSALWCRCYFGTTEGAGELAPND 416	Db	365 PKIGTVARLAODGSNRQPKFIIPSTADRLRGERDVGSLVLSALWCRCYFAGKSDSG--- 421
Qy	428 PTYGDAEWKLAKADDFESSLKLPADFGRWDLDT--SELDQKVIVLR-----KIIRKGV 479	Qy	424 ESSEPTYGDAE-----WKLAKADDFESSLKLPADFGRWDLDTSELDQKVIV-LKKIIR 476
Db	417 PNW--ERLVAVA--QAQKDRPA--AWLEMDIYAEGVGRNPVAVIETSETTALKAIWARGS 468	Db	422 --KEIVFNDANADRLHAAAVAAKDDPMFALSL--DIFGDVAQSDLFRFRFAHALKVLWE 477
Qy	480 KAAI 483	Qy	477 KGVKAAI 483
Db	469 RAVI 472	Db	478 KQTRATL 484
<p>RESULT 10 Q92N07 PRELIMINARY; PRT; 494 AA.</p>		<p>RESULT 11 Q8Z6Z6 PRELIMINARY; PRT; 488 AA.</p>	
ID	Q92N07	ID	Q8Z6Z6
AC	Q92N07	AC	Q8Z6Z6
DT	01-DEC-2001 (Tremblrel. 19, Created)	DT	01-MAR-2002 (Tremblrel. 20, Created)
DT	01-DEC-2001 (Tremblrel. 19, Last sequence update)	DT	01-MAR-2002 (Tremblrel. 20, Last sequence update)
DT	01-MAR-2002 (Tremblrel. 20, Last annotation update)	DT	01-JUN-2002 (Tremblrel. 21, Last annotation update)
DE	Probable mannitol 2-dehydrogenase protein (EC 1.1.1.67).	DE	Putative D-mannanate oxidoreductase (EC 1.1.1.57).
GN	MTLK OR R02440 OR SMC01501	GN	STY1553
OS	Rhizobium meliloti (Sinorhizobium meliloti).	OS	Salmonella typhi.
OC	Bacteria; Proteobacteria; alpha subdivision; Rhizobiaceae group;	OC	Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
OC	Rhizobiaceae; Sinorhizobium.	OC	Salmonella.
OX	NCBI_TaxID=382;	OX	NCBI_TaxID=601;
RN	[1]	RN	[1]
RP	SEQUENCE FROM N.A.	RP	SEQUENCE FROM N.A.
RC	STRAIN=1021;	RC	STRAIN=CT18;
RX	MEDLINE=21396507; PubMed=11481430;	RX	MEDLINE=21534947; PubMed=11677608;
RA	Capela D., Barloy-Hubler F., Gouzy J., Bothe G., Ampe F., Batut J.,	RA	Parkhill J., Dougan G., James K.D., Thomson N.R., Pickard D., Wain J.,
RA	Boistard P., Becker A., Boutry M., Cadieu E., Dreano S., Gloux S.,	RA	Churcher C., Mungall K.L., Bentley S.D., Holden M.T.G., Sebahia M.,
RA	Godrie T., Goffeau A., Kahn D., Kiss E., Lelaure V., Masuy D.,	RA	Baker S., Basham D., Brooks K., Chillingworth T., Connor P.,
RA	Pohl T., Portetelle D., Puehler A., Purnelle B., Rampsperger U.,	RA	Cronin A., Davis P., Davies R.M., Dowd L., White N., Farrar J.,
RA	Renard C., Thebault P., Vandenbol M., Weidner S., Galibert F.;	RA	Feltwell T., Hamlin N., Haque A., Hien T.T., Holroyd S., Jagels K.,

OS Salmonella typhi.
OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
OC Salmonella.
OX NCBI_TaxID=602, 601;
RN [1]
RP SEQUENCE FROM N.A.
RC SPECIES=S.typhimurium; STRAIN=LT2 / SGSCI412 / ATCC 700720;
RX MEDLINE=21534948; PubMed=11677609;
RA McClelland M., Sanderson K.E., Spieth J., Clifton S.W., Latreille P.,
RA Courtney L., Porwollik S., Ali J., Dante M., Du F., Hou S., Layman D.,
RA Leonard S., Nguyen C., Scott K., Holmes A., Grewal N., Mulvaney E.,
RA Ryan E., Sun H., Florea L., Miller W., Stoneking T., Nhan M.,
RA Waterston R., Wilson R.K.;
RT "Complete genome sequence of Salmonella enterica serovar Typhimurium
LT2.";
RL Nature 413:852-856(2001).
RN [2]
RP SEQUENCE FROM N.A.
RC SPECIES=S.typhi; STRAIN=CT18;
RX MEDLINE=21534947; PubMed=11677608;
RA Parkhill J., Dougan G., James K.D., Thomson N.R., Pickard D., Main J.,
RA Churcher C., Mungall K.L., Bentley S.D., Holden M.T.G., Sebaihia M.,
RA Baker S., Basham D., Brooks K., Chillingworth T., Connor P.,
RA Cronin A., Davis P., Davies R.M., Dowd L., White N., Farrar J.,
RA Feltwell T., Hamlin N., Haque A., Hien T.T., Holroyd S., Jagers K.,
RA Krogh A., Larsen T.S., Leather S., Moule S., O'Gaora P., Parry C.,
RA Quail M., Rutherford K., Simmonds M., Skelton J., Stevens K.,
RA Whitehead S., Barrrell B.G.;
RT "Complete genome sequence of a multiple drug resistant Salmonella
enterica serovar Typhi CT18.";
RL Nature 413:848-852(2001).
DR EMBL; AE008844; AAL22010.1; -;
DR EMBL; AL627277; CAD02968.1; -;
DR InterPro; IPR000669; Mannitol_dh.
DR InterPro; IPR001993; Mitoch_carrier.
DR Pfam; PF01232; Mannitol_dh; 1;
DR PRINTS; PR00084; MTLDRGNASE.
DR PROSITE; PS00974; MANNITOL_DHGENASE; 1.
DR PROSITE; PS00215; MITOCH_CARRIER; UNKNOWN_1.
KW Oxidoreductase; Hypothetical protein; Complete proteome.
SQ SEQUENCE 490 AA; 54007 MW; A521610BB383A9FF CRC64;

Query Match 29.0%; Score 730; DB 16; Length 490;
Best Local Similarity 36.5%; Pred. No. 1.9e-44;
Matches 156; Conservative 83; Mismatches 182; Indels 6; Gaps 5;

QY 5 ETLKSLPANVQAPPYDIDGKPGIVHFGNFFRAHEAFYVEQILEHA-PDWAIVGVGL- 62
DB 3 QNIATAQSVARPNDWKSRLSVRIVHLGCGAFRAHQALFTHHLEKSDSDWGICEVNL 62
QY 63 TGSRSKKAEEFKAQDCILYSLETAPSGKSTVRVMGALRDYLLAPADPEA-VLKHLVDP 121
DB 63 PCND--ARLIANLKAQNLYTVAERG-AESTELKIIGSMKEALHPEFDGHAGILAAMARP 119
QY 122 ATRIVSWTTTEGYNINETTGAFDLENAAVKADLNKPEKSPVFGYVVEALRRRWDAQK 181
DB 120.ETAIVSLVTERKGYCTDPASGELDVNNPLIQNDLAHPQOPKSAIGYIVEALNMRREQGLK 179
QY 182 AFTVNSCDNLRLNGVAKAFGLYKAKARPELAKWIEENATPPNGMVDRIPTVSAEIAK 241
DB 180 AFTVLSCDNVRENGHVAKAANLGLAKARDAALAAIADNVTFCTWVDRIVPAAATEETLQ 239
QY 242 KLNAAAGLDDDLPLVAEDFHOWLEDOFADGRPPLEKAGQVMGVQVTDWEYVKIRMLNAG 301
DB 240 LVADQLGVYDPCAIACEPPFQWVIEDNFVNGRPDMDTVGAQFVADVVPFEMKRLNCS 299
QY 302 HVMLCFPGILVGYENVDDAIEDSELLGNLKNYLNKDVIPTLKAPSGMTLEGYRDSVISRF 361
DB 300 HSFLAYLGYLGGYDITADTMTNPAYRRAALALMDLDEQAPTLSPMEGTDLGYANLLIARF 359
QY 362 SNKAMSDOTLRATSDGCKSVQVFWETVETPRATEDKRLSLRIAFGLASYLEMLRGDEKGG 421
DB 360 TNPULKHTWQIAMDGSQKLPRLLDPLVRLHLQGGDDYRRLTLGVAGWMRYVGGVDEQCK 419

QY 422 TYESSEP 428
DB 420 TIDVVDP 426
Search completed: May 8, 2003, 16:33:38
Job time : 41 secs

THIS PAGE BLANK (USPTO)

GenCore version 5.1.6
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM nucleic - nucleic search, using sw model
Run on: May 19, 2003, 00:18:20 ; Search time 10357 Seconds
(without alignments)
11563.005 Million cell updates/sec
Title: US-09-926-163B-1
Perfect score: 4115
Sequence: 1 aagcttgatcgtgcaggt.....tccggcgccctgaagctt 4115
Scoring table: OLIGO_NUC
Gapex 60.0 , Gapext 60.0
Searched: 2054640 seqs, 14551402878 residues
Word size : 0
Total number of hits satisfying chosen parameters: 4109280
Minimum DB seq length: 0
Maximum DB seq length: 2000000000
Post-processing: Listing first 45 summaries

Database : GenEmbl.*
1: gb_ba.*
2: gb_hcg.*
3: gb_in.*
4: gb_om.*
5: gb_ov.*
6: gb_pat.*
7: gb_ph.*
8: gb_pi.*
9: gb_pr.*
10: gb_ro.*
11: gb_sts.*
12: gb_sy.*
13: gb_un.*
14: gb_vi.*
15: em_ba.*
16: em_fun.*
17: em_hum.*
18: em_in.*
19: em_mu.*
20: em_om.*
21: em_or.*
22: em_ov.*
23: em_pat.*
24: em_ph.*
25: em_pl.*
26: em_ro.*
27: em_sts.*
28: em_un.*
29: em_vi.*
30: em_htg_hum.*
31: em_htg_inv.*
32: em_htg_other.*
33: em_htg_mus.*
34: em_htg_pln.*
35: em_htg_rod.*
36: em_htg_mam.*
37: em_htg_vrt.*
38: em_sy.*
39: em_htgo_hum.*
40: em_htgo_mus.*
41: em_htgo_other.*

Pred. No. is the number of results predicted by chance to have a

score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description
1	4073	99.0	4073	1	AB028937	AB028937 Gluconoba
2	40	1.0	4643	12	AF027127	AF027127 Eukaryoti
3	38	0.9	91	12	SYNIIFOKI	M30778 MboII/FOKI
4	38	0.9	360	6	AX256425	AX256425 Sequence
5	38	0.9	416	5	DRE238981	AJ238981 Danio rer
6	38	0.9	1195	6	I09132	I09132 Sequence 1
7	38	0.9	2442	1	MTGLMS	AJ000333 Mycobacte
8	38	0.9	4404	12	SYNCATBLB	M80483 Cloning vec
9	38	0.9	4496	12	BLCAT2DNA	X64410 plasmid pBL
10	38	0.9	4711	12	AF027128	AF027128 Eukaryoti
11	38	0.9	7055	12	AF102233	AF102233 Transposo
12	38	0.9	8142	12	AF076213	AF076213 Promoter
13	38	0.9	10737	12	XXU02428	U02428 Cloning vec
14	38	0.9	10850	12	U02455	U02455 Cloning vec
15	37	0.9	92	6	A25917	A25917 Expression
16	37	0.9	92	6	A25918	A25918 Expression
17	37	0.9	341	11	G49583	G49583 SHGC-60412
18	37	0.9	586	6	AR007549	AR007549 Sequence
19	37	0.9	586	6	AR029479	AR029479 Sequence
20	37	0.9	586	6	I90252	I90252 Sequence 10
21	37	0.9	916	11	G63570	G63570 SHGC-141568
22	37	0.9	2907	12	XXU13871	U13871 pT73D clon
23	37	0.9	3796	6	E49323	E49323 Infectious
24	37	0.9	4674	12	XXU02437	U02437 Cloning vec
25	37	0.9	4910	6	AR027100	AR027100 Sequence
26	37	0.9	4910	6	AR036062	AR036062 Sequence
27	37	0.9	4910	6	AR160052	AR160052 Sequence
28	37	0.9	4910	6	AR209764	AR209764 Sequence
29	37	0.9	5027	12	AF276982	AF276982 Integrati
30	37	0.9	5321	12	AF129072	AF129072 Cloning v
31	37	0.9	6282	12	AF069723	AF069723 Expressio
32	37	0.9	6534	12	AF069724	AF069724 Expressio
33	37	0.9	6541	12	XXU02433	U02433 Cloning vec
34	37	0.9	6894	12	XXU02435	U02435 Cloning vec
35	37	0.9	7036	12	AF069722	AF069722 Expressio
36	37	0.9	7069	12	U02442	U02442 Cloning vec
37	37	0.9	7164	6	AX195206	AX195206 Sequence
38	37	0.9	7164	12	U02451	U02451 Cloning vec
39	37	0.9	7481	12	XXU02438	U02438 Cloning vec
40	37	0.9	7522	12	AF069717	AF069717 Expressio
41	37	0.9	7612	6	AX009487	AX009487 Sequence
42	37	0.9	7897	6	AR093383	AR093383 Sequence
43	37	0.9	7897	6	AR142583	AR142583 Sequence
44	37	0.9	8540	6	AR198722	AR198722 Sequence
45	37	0.9	8932	6	AR069417	AR069417 Sequence

ALIGNMENTS

RESULT 1
AB028937
LOCUS
DEFINITION
AB028937
Gluconobacter oxydans gene for sorbitol dehydrogenase, complete
cds.
ACCESSION
AB028937.1 GI:9049448
VERSION
sorbitol dehydrogenase.
KEYWORDS
Gluconobacter oxydans (strain:G624) DNA.
SOURCE
Gluconobacter oxydans
ORGANISM
Bacteria; Proteobacteria; alpha subdivision; Acetobacteraceae;
Gluconobacter.
REFERENCE
1 (sites)
Shibata,T., Ichikawa,C., Matsuura,M., Takata,Y., Noguchi,Y.,
Saito,Y. and Yamashita,M.

TITLE		Cloning of a gene for D-Sorbitol dehydrogenase from <i>Gluconobacter oxydans</i> G624 and expression of the gene in <i>Pseudomonas putida</i> IF03738	
JOURNAL REFERENCE		J. Biosci. Bioeng. 89, 463-468 (2000)	
AUTHORS		Shibata, T. and Saito, Y.	
TITLE		Direct Submission	
JOURNAL		Submitted (15-JUN-1999) Takashi Shibata, Fujisawa Pharmaceutical Co., Ltd., Fermentation Development Laboratories; Shinkawa-cho, Nakagawara 156, Nishikasugai-gun, Aichi 452-0915, Japan (E-mail: takashi.shibata@po.fujisawa.co.jp, Tel: 81-52-401-0275, Fax: 81-52-400-1380)	
FEATURES		Location/Qualifiers	
source		1. .4073	
		/organism="Gluconobacter oxydans"	
		/strain="G624"	
		/db_xref="taxon:442"	
CDS		501..1958	
		/codon_start=1	
		/transl_table=11	
		/product="sorbitol dehydrogenase"	
		/protein_id="BAA99414.1"	
		/db_xref="GI:9049449"	
		/translation="MITRETIKSLPANVQAPPYDIDGIKPGIVHFGVGNFFRAHEAFYVQIIEHPADVAIVGGLTGDGRSKKAEFAKQDCLYSLTETAPSKSTVRYVMGALRDYLLAPADVAIVLKHLPDIRIVSMITTEGGYNINLETGAFDLENAAVKDLNPKP PSTVGYVVEALRRWRDAGGAKFTVMSCDNLRHNGVNARKAFGLYKARDELPKAWIE ENATFPNGVDVITPVSAETAKNLNRAAGLDDPLVAEDFHOWVLEDFADGRPPL ERKATVGLKAPSGMTLEGYRDSVTSRFSNKAMSQDTLRASDSDGSKVQVFWETVRRAL IDKRDLSRIAFGLASYILEMLGRDKEGTYESSEPTYGDAEMKLAKADDPFESSLKLP AFDGWRDLDTSELDOKVIVLRKIIIEKGVKAIPA"	
BASE COUNT		823 a 1122 c 1259 g 869 t	
ORIGIN			
Query Match		99.0%; Score 4073; DB 1; Length 4073;	
Best Local Similarity		100.0%; Pred. No. 0;	
Matches 4073; Conservative		0; Mismatches 0; Indels 0; Gaps 0;	
Qy	37	GGTTTTGGCAGCGCTCCCTAGATTGATCGCGCGTCTGTGACCGACATGATCGTGGTGC	96
Db	1	GGTTTTGGCAGCGCTCCCTAGATTGATCGCGCGTCTGTGACCGACATGATCGTGGTGC	60
Qy	97	ACGTGCCATTGGCAGCGGCGTGCACCGGGAAACACAGCGCTGTGTCCTTTGTACAAAGG	156
Db	61	ACGTGCCATTGGCAGCGGCGTGCACCGGGAAACACAGCGCTGTGTCCTTTGTACAAAGG	120
Qy	157	GCTAGTCATCGCTCGCTGGCTGGCACAATAGTTGCCAAGACAGTATGCCGCGCAAGCA	216
Db	121	GCTAGTCATCGCTCGCTGGCTGGCACAATAGTTGCCAAGACAGTATGCCGCGCAAGCA	180
Qy	217	GAACACAGTAAACAGCAGTCCGAGACAGAGAAATCCTCGCGCTCCTACCGCCATTGGA	276
Db	181	GAACACAGTAAACAGCAGTCCGAGACAGAGAAATCCTCGCGCTCCTACCGCCATTGGA	240
Qy	277	AGACGAGACCGTCTGAGATCGCTTTGTGATGTCCTGTTCGCGGAGGATCTCGAAGC	336
Db	241	AGACGAGACCGTCTGAGATCGCTTTGTGATGTCCTGTTCGCGGAGGATCTCGAAGC	300
Qy	337	GGCTGTTGGGGTCTCATCGCTTCTGATCGGAGTCCGCAAAAGTCTCTGAACAGGACG	396
Db	301	GGCTGTTGGGGTCTCATCGCTTCTGATCGGAGTCCGCAAAAGTCTCTGAACAGGACG	360
Qy	397	TCCCGCGGAGGCGATCAGAGTGGAAATGGCTCTCTGTTGAAACCGTCAATTCGGTTTTTA	456
Db	361	TCCCGCGGAGGCGATCAGAGTGGAAATGGCTCTCTGTTGAAACCGTCAATTCGGTTTTTA	420
Qy	457	CGTTGTTTCGGGCTATGATGCACATGCCCGGCGCTTGTGCTGCCCGCTCAGCGACGCG	516
Db	421	CGTTGTTTCGGGCTATGATGCACATGCCCGGCGCTTGTGCTGCCCGCTCAGCGACGCG	480
Qy	517	CGGAAACACCGAGAGATTCATGATTACGCGGCAACCCCTTAAGTCTCTCTCTGCCAATG	576

1657 CTAGCGATGGGTGTTTCCAAAGTTTCAGGTGTTCTTGAACGAAACCGTGGTGGCGGATCG 1716
1621 CTAGCGATGGGTGTTTCCAAAGTTTCAGGTGTTCTTGAACGAAACCGTGGTGGCGGATCG 1680
1717 AAGACAAGCGGAGCTGTACAGTATAGGTTTCGGAATTCATCTCAAAATGCTGC 1776
1681 AAGACAAGCGGAGCTGTACAGTATAGGTTTCGGAATTCATCTCAAAATGCTGC 1740
1777 GTGGTTCGGACGAGAAGCGGGAGCTATGAATCGTCCGAGCCGACTTATGGCGACGCG 1836
1741 GTGGTTCGGACGAGAAGCGGGAGCTATGAATCGTCCGAGCCGACTTATGGCGACGCG 1800
1837 AATGGAAGTTGGCAAGCGGAGCAGCTTCGAAAGCTCTCTGAAGCTCCCGCGCTTCGATG 1896
1801 AATGGAAGTTGGCAAGCGGAGCAGCTTCGAAAGCTCTCTGAAGCTCCCGCGCTTCGATG 1860
1897 GTGGCGCGATCTGGATACGTCGCAACTGGATCAAAAGTGTATCGTCTCGGAAGATCA 1956
1861 GTGGCGCGATCTGGATACGTCGCAACTGGATCAAAAGTGTATCGTCTCGGAAGATCA 1920
1957 TCCGCGAAAAGGGGTAAAGCCGCAATCCCGGCTCGAATTCGGCTTTTAGGGTAGCGAC 2016
1921 TCCGCGAAAAGGGGTAAAGCCGCAATCCCGGCTCGAATTCGGCTTTTAGGGTAGCGAC 1980
2017 TGAACAGAAAACCGCGCTCTGGAAGGAGCGCGGTTTTTTTTTATGCTCAGATCTGTCCCA 2076
1981 TGAACAGAAAACCGCGCTCTGGAAGGAGCGCGGTTTTTTTTTATGCTCAGATCTGTCCCA 2040
2077 TCAGGACAAGGATCACGACGACACGATCAGGACAGTCCGCTCGAGGGGAGCCCATTT 2136
2041 TCAGGACAAGGATCACGACGACACGATCAGGACAGTCCGCTCGAGGGGAGCCCATTT 2100
2137 TCGAACGTGTACGGCATGACGGCAGCGACCGAGATCAGGATTAACAAGAGATCAGTCC 2196
2101 TCGAACGTGTACGGCATGACGGCAGCGACCGAGATCAGGATTAACAAGAGATCAGTCC 2160
2197 CATGGCACATCTCTTGGCCGGTTGAGACTGGTCTGTGTTCCGGGTGTCTAAAAAGTTTC 2256
2161 CATGGCACATCTCTTGGCCGGTTGAGACTGGTCTGTGTTCCGGGTGTCTAAAAAGTTTC 2220
2257 CGTAGGGCGGAAAGATCAAGCTGTGCGTTCGCGCTTAATCCGCTCCAGCGCGCATTTG 2316
2221 CGTAGGGCGGAAAGATCAAGCTGTGCGTTCGCGCTTAATCCGCTCCAGCGCGCATTTG 2280
2317 ATGCGGGCCACCCGGTCTGTGCGGTTGCGCTCTGTCTCTGACATAGGTTTCTGGGCG 2376
2281 ATGCGGGCCACCCGGTCTGTGCGGTTGCGCTCTGTCTCTGACATAGGTTTCTGGGCG 2340
2377 AGCAGTCCGGATGATGTTCCGGATCAGGTTGCGGACGCGCAGCGGATTTCTGTGTCA 2436
2341 AGCAGTCCGGATGATGTTCCGGATCAGGTTGCGGACGCGCAGCGGATTTCTGTGTCA 2400
2437 GTTCGCTGCGGGTATGTCGCGAGATACGATAGGATCCGGCTGTTCCGCTGGCGCG 2496
2401 GTTCGCTGCGGGTATGTCGCGAGATACGATAGGATCCGGCTGTTCCGCTGGCGCG 2460
2497 CGATTGTTGCGGCTTTTCGGCCGGTCCCATGCTCTGCGGCGAGCGCAAAATGCCCCGTGA 2556
2461 CGATTGTTGCGGCTTTTCGGCCGGTCCCATGCTCTGCGGCGAGCGCAAAATGCCCCGTGA 2520
2557 ACGGCTGCGAATAATCGATTTCTTCGGGTGAAAGTTCGGGGTGGGCGCGGATCGGCA 2616
2521 ACGGCTGCGAATAATCGATTTCTTCGGGTGAAAGTTCGGGGTGGGCGCGGATCGGCA 2580
2617 CGGCGGATACGGAACAGTCCGCTCATGAGTTCTCAAGCGCGCGCTTATATCGGCGATAG 2676
2581 CGGCGGATACGGAACAGTCCGCTCATGAGTTCTCAAGCGCGCGCTTATATCGGCGATAG 2640
2677 GCCTTGGCCATTTTCGGGGGATACATCTCGAAATCGTCCGCTCGCGGGCGCGATCG 2736
2641 GCCTTGGCCATTTTCGGGGGATACATCTCGAAATCGTCCGCTCGCGGGCGCGATCG 2700

2737 AACAGCATGCCGACTTCTCTGGTGTATTATCGGGGGGAACTGGAAGCAGGTCTTGAAGCG 2796
2701 AACAGCATGCCGACTTCTCTGGTGTATTATCGGGGGGAACTGGAAGCAGGTCTTGAAGCG 2760
2797 TTGATTTTCGTTCTCGGTTACCGGGCCGCTCGATCTTCGCCAGCTTCGCCACAGGGCAACA 2856
2761 TTGATTTTCGTTCTCGGTTACCGGGCCGCTCGATCTTCGCCAGCTTCGCCACAGGGCAACA 2820
2857 AGGCGCATGGCGTAAAGCTGATCTCTGTTTCCAGGGCCGCGAGCAATCTTGGCAGCGCG 2916
2821 AGGCGCATGGCGTAAAGCTGATCTCTGTTTCCAGGGCCGCGAGCAATCTTGGCAGCGCG 2880
2917 AAAAAGCGCGCGCTGTGGGATTCGGGATTCGGGCAATTCGGGGAAAGCGCTCACTCCAGCG 2976
2881 AAAAAGCGCGCGCTGTGGGATTCGGGATTCGGGCAATTCGGGGAAAGCGCTCACTCCAGCG 2940
2977 CCCGTTGAGGCTTTGAGTAGGCAACCGTTATTCGGGGCATGCCAGCGCTTCGCCCATC 3036
2941 CCCGTTGAGGCTTTGAGTAGGCAACCGTTATTCGGGGCATGCCAGCGCTTCGCCCATC 3000
3037 AGTCTCCGAAAGGACCAACCGCGAAGCCCGACACCCAGCAACATCTTGCCCCAG 3096
3001 AGTCTCCGAAAGGACCAACCGCGAAGCCCGACACCCAGCAACATCTTGCCCCAG 3060
3097 ATAGCCATGTCTCAACCTTAGCACGCCGCTCAGAGGGCAAAATGACAGATCGCAGGCTA 3156
3061 ATAGCCATGTCTCAACCTTAGCACGCCGCTCAGAGGGCAAAATGACAGATCGCAGGCTA 3120
3157 GGTCTAGGTGCTGATCGCCAAACCGCGGCTTCGGGTGCTAGAGCTAGGAGTTAC 3216
3121 GGTCTAGGTGCTGATCGCCAAACCGCGGCTTCGGGTGCTAGAGCTAGGAGTTAC 3180
3217 GAACCTTATCGCTGCTCATGCTTTTGAGGCGCAGGTTCTCTGTTCGTTTCATACCGAT 3276
3181 GAACCTTATCGCTGCTCATGCTTTTGAGGCGCAGGTTCTCTGTTCGTTTCATACCGAT 3240
3277 ATTTTATGCCCCACTTGATCCAGACTGCTACTTCGATCCCTTCGCGCTCTGATGACGAA 3336
3241 ATTTTATGCCCCACTTGATCCAGACTGCTACTTCGATCCCTTCGCGCTCTGATGACGAA 3300
3337 CTGATGATCTTTTGATCAAGCGTCTCCAATGCGGTGCGAAGTCTCAACTGSGTTG 3396
3301 CTGATGATCTTTTGATCAAGCGTCTCCAATGCGGTGCGAAGTCTCAACTGSGTTG 3360
3397 CGGAAGCGGATCAATAATGGTTTCGGATTCGGCGGGCGCTGCTGTTTCATCTCGGGCG 3456
3361 CGGAAGCGGATCAATAATGGTTTCGGATTCGGCGGGCGCTGCTGTTTCATCTCGGGCG 3420
3457 GTTCTGTCCATCTCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTT 3516
3421 GTTCTGTCCATCTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTT 3480
3517 GCGCAGGATATTCGCTGCTTCCGCTGCTTCCAGGGCCGCTCTTCCGCTGCTGCTGCTGCT 3576
3481 GCGCAGGATATTCGCTGCTTCCGCTGCTTCCAGGGCCGCTCTTCCGCTGCTGCTGCTGCT 3540
3577 CAACATCCGATTTGCTGCTGCTTCCGCGGAAAAAGCGGAGAACTAACCTTCTGCTG 3636
3541 CAACATCCGATTTGCTGCTGCTTCCGCGGAAAAAGCGGAGAACTAACCTTCTGCTG 3600
3637 ACGTGTTCGTAAGATGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 3696
3601 ACGTGTTCGTAAGATGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 3660
3697 GGTGCTCGCGGATCGAGAGAACCCAGACAGCAAGCTCTGCTGCGGACGCTGCTGCTGCT 3756
3661 GGTGCTCGCGGATCGAGAGAACCCAGACAGCAAGCTCTGCTGCGGACGCTGCTGCTGCT 3720
3757 GCCATCGCTCCAGTATAGCCAGAACTCGGGTTCCAGTTCGCGGACGCTCCGCTGCTCT 3816
3721 GCCATCGCTCCAGTATAGCCAGAACTCGGGTTCCAGTTCGCGGACGCTCCGCTGCTCT 3780
3817 GACAGACAGGCTGCGCTTGTGACGAGATCACTCATCTCCGGTTGTTTCTCAAGGCGCTTCA 3876

|||||
Db 3781 GACAGACAGCGCTCGTTTTCAGGAGATCACTATTCGGTGTGTTCTCAAGCGCTTCA 3840
QY 3877 AAGCCCATTTGCGGTTTCGGAACATCAGGTCGGATCACTACGAGCTCCCGCGCAG 3936
Db 3841 AAGCCCATTTGCGGTTTCGGAACATCAGGTCGGATCACTACGAGCTCCCGCGCAG 3900
QY 3937 AAGATATAAGCGACGATCGCGCGGTTTCGCGATCGGATCAGGACATTTACGTACGAAC 3996
Db 3901 AAGATATAAGCGACGATCGCGCGGTTTCGCGATCGGATCAGGACATTTACGTACGAAC 3960
QY 3997 CGGTTGCGTCCCAATCCGTTTACCGCGAGAGCCAGAAAAAAGCTCCGGAATGTCCGCAATTA 4056
Db 3961 CGGTTGCGTCCCAATCCGTTTACCGCGAGAGCCAGAAAAAAGCTCCGGAATGTCCGCAATTA 4020
QY 4057 TCCAGCCGACGAGTTCGTCGAGTTTTCGTCGATCAGCTCCGGCGGCGGCTG 4109
Db 4021 TCCAGCCGACGAGTTCGTCGAGTTTTCGTCGATCAGCTCCGGCGGCGGCTG 4073

RESULT 2
AF027127
LOCUS AF027127 4643 bp DNA circular SYN 08-NOV-1997
DEFINITION Eukaryotic luciferase expression vector pTALUC+, complete
ACCESSION AF027127
VERSION AF027127.1 GI:2598100
KEYWORDS Expression vector pTALUC+.
SOURCE Expression vector pTALUC+.
ORGANISM Artificial sequences; vectors.
REFERENCE 1 (bases 1 to 4643)
AUTHORS Altschmid, J. and Duschl, J.
TITLE Set of optimized luciferase reporter gene plasmids compatible with
widely used CAT vectors
JOURNAL BioTechniques 23 (3), 436-438 (1997)
MEDLINE 9298214
PUBMED 9298214
REFERENCE 2 (bases 1 to 4643)
AUTHORS Altschmid, J. and Duschl, J.
TITLE Direct Submission
JOURNAL Submitted (26-SEP-1997) Physiological Chemistry I/Biochemistry,
University of Wuerzburg, Am Hubland, Wuerzburg 97074, Germany
FEATURES
Location/Qualifiers
1..4643
/organism="Expression vector pTALUC+"
/db_xref="taxon:68900"
/plasmid="pTALUC+"

misc_feature 1..36
/note="polylinker"
misc_feature 37..131
/note="tk-promoter"
gene 170..1822
/gene="luc+"
CDS 170..1822
/gene="luc+"
/codon_start=1
/product="luciferase"
/protein_id="AAB83989.1"
/db_xref="GI:2598101"
/translation="MEDAKNKKGPAPYPLEDGTAGQLHKMKRYALVPGTIAFTD
AHIEVDITYAEYFENSLAEAMKRYGLNTNHRIVVCSNSLOFFMPVLGALFIGVAV
APANDIYNERELLNSMGIQSOPVVFVSKKGLQILNOKKLPILQKIIIMDSKTDYOG
FSMYFTVTHLPGFENYDFVPESEFDKFTIALIMNSGSGTGLPKGVALPHRTACVR
FOSHADPLFGNQLIPDPTAILSVPPHFGCGMTTGLYLICGFRVVLVYRFEELFLRS
LQDYKIQSALLVPLTFSFAKSTLIDKLDNLHEIASGAPLAKSKEGVAEYAKRPHLP
RGIMGYTETSITAILITPEGDDPKPAGVGVVPEFAKVDLDGTGLTGVNORGEVCY
RPMGMSYVNNPATLIDKGLHSGDIAYWDEDEHFFIVDLKSLIKYKGYQVA
PAELSLILQHPNIFDAGVAGLPDDAGELPAVAVVLEHGKMTKEIVDVVASQVTT
AKKLGGVVVDEVPKGLTGKLDARKITREILIKAKKGKGIIV"
1854..2075
/note="splice and polyadenylation signals"

misc_feature 2078..2094

/note="polylinker"
complement(3277..4137)
/gene="bla"
complement(3277..4137)
/gene="bla"
/codon_start=1
/product="beta-lactamase"
/protein_id="AAB83990.1"
/db_xref="GI:2598102"
/translation="MSIQHERVALIPFAAFCLPVAHPETLVYVKDAEDQLGARVGY
IEDLSNGKILESFRPEPPEPMSTFKVLCGAVLSRIDAGQEQLGRIIRHYSDNLVE
YSPYTERKLDGTVTRVRELCSAAITMSDNTAANLLITIGGPKELTAFLNHMGDHTRL
DRWPELNEAIPNDERDTTPVAMATTLRKLLTGLLTLASROQLIDMEADKVGPL
LRSLALPAGWFIADKSGAGERSGIIAALGPDGKPSRVIVYITGTSQATWDERNRQIA
EIGSLALIKHW"
misc_feature 4269..4640
/note="polyadenylation signals"
BASE COUNT 1243 a 1078 c 1097 g 1225 t
ORIGIN

Query Match 1.0%; Score 40; DB 12; Length 4643;
Best Local Similarity 100.0%; Pred. No. 9.3e-09;
Matches 40; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 AAGCTTCGATCGCTCGAGTTCGACTCTAGAGATCCGGTT 40
|||||
Db 1 AAGCTTCGATCGCTCGAGTTCGACTCTAGAGATCCGGTT 40

RESULT 3
LOCUS SYNIFOKI 91 bp DNA linear SYN 27-APR-1993
DEFINITION MboII/FokI multiple cloning site cassette DNA in plasmid pRBG177.
ACCESSION M30778
VERSION M30778.1 GI:208631
KEYWORDS Synthetic DNA.
SOURCE synthetic construct
ORGANISM artificial sequences.
REFERENCE 1 (bases 1 to 91)
AUTHORS Hasan, N., Kur, J. and Szybalski, W.
TITLE An MboII/FokI trimming plasmid allowing consecutive cycles of
precise 1- to 12-base-pair deletions in cloned DNA
JOURNAL Gene 82 (2), 305-311 (1989)
MEDLINE 90060819
PUBMED 2583524
FEATURES Location/Qualifiers
source
1..91
/organism="synthetic construct"
/db_xref="taxon:32630"

BASE COUNT 19 a 26 c 25 g 21 t
ORIGIN

Query Match 0.9%; Score 38; DB 12; Length 91;
Best Local Similarity 100.0%; Pred. No. 1.2e-07;
Matches 38; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 AAGCTTCGATCGCTCGAGTTCGACTCTAGAGATCCGG 38
|||||
Db 91 AAGCTTCGATCGCTCGAGTTCGACTCTAGAGATCCGG 54

RESULT 4
LOCUS AX256425 360 bp DNA linear PAT 10-OCT-2001
DEFINITION Sequence 9 from Patent WO0076523.
ACCESSION AX256425
VERSION AX256425.1 GI:16075243
KEYWORDS synthetic construct.
SOURCE synthetic construct.
ORGANISM artificial sequences.
REFERENCE 1 (bases 1 to 360)

AUTHORS Forman,B.M., Beard,R.L. and Chandraratna,R.A.
 TITLE Methods for modulating fxr receptor activity
 JOURNAL Patent: WO 0076523-A 9 21-DEC-2000;
 Allergan Sales, Inc. (US)

FEATURES
 source
 1..360
 /organism="synthetic construct"
 /db_xref="taxon:32630"
 /note="Promoter region of plasmid TK-luc"

BASE COUNT 88 a 103 c 103 g 66 t

ORIGIN
 Query Match 0.9%; Score 38; DB 6; Length 360;
 Best Local Similarity 100.0%; Pred. No. 1.1e-07;
 Matches 38; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 AAGCTTGATCGCTGCAGGTGACTCTAGAGGATCCGG 38
 ||||||||||||||||||||||||||||||||||||||||
 Db 42 AAGCTTGATCGCTGCAGGTGACTCTAGAGGATCCGG 79

RESULT 5
 DRE238981 416 bp mRNA linear VRT 06-JAN-2000
 LOCUS Danio rerio mRNA for activin beta B protein, partial.
 DEFINITION
 ACCESSION AJ238981
 VERSION AJ238981.1 GI:4867810
 KEYWORDS activin beta B protein.
 SOURCE Danio rerio.
 ORGANISM Danio rerio
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Actinopterygii; Neopterygii; Teleostei; Ostariophysi;
 Cypriniformes; Cyprinidae; Danio.

REFERENCE 1 (bases 1 to 416)
 AUTHORS Rodaway,A., Takeda,H., Koshida,S., Broadbent,J., Price,B.,
 Smith,J.C., Patient,R. and Holder,N.
 TITLE Induction of the mesoderm in the zebrafish germ ring by yolk
 cell-derived TGF-beta family signals and discrimination of mesoderm
 and endoderm by FGF
 JOURNAL Development 126 (14), 3067-3078 (1999)
 MEDLINE 99307072
 PUBMED 10375499
 REFERENCE 2 (bases 1 to 416)
 AUTHORS Smith,J.C.
 TITLE Direct Submission
 JOURNAL Submitted (14-MAY-1999) Smith J.C., Developmental Biology, National
 Institute for Medical Research, The Ridgeway, Mill Hill, London NW7
 1AA, UNITED KINGDOM

FEATURES
 source
 1..416
 /organism="Danio rerio"
 /db_xref="taxon:7955"
 1..416
 /gene="activin beta B"
 <1..>416
 /gene="activin beta B"
 /function="growth factor"
 /codon_start=1
 /product="activin beta B protein"
 /protein_id="CAB43092.1"
 /db_xref="GI:4867811"
 /db_xref="SPTREMBL:Q9W679"
 /translation="GLEQLKACIQVDSRGSGFCEDGNNGLCRCROOFYIDFRILIGNW
 DWIIAPGYGNYGEGPCAYMGVPGSASSFTAVVNOYRMGMSPGSVNSCCIPTK
 LSTWSMLVFDDEYNIVKRDVPMVEGCGEFDI"
 BASE COUNT 97 a 107 c 113 g 99 t

Query Match 0.9%; Score 38; DB 5; Length 416;
 Best Local Similarity 100.0%; Pred. No. 1.1e-07;
 Matches 38; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 AAGCTTGATCGCTGCAGGTGACTCTAGAGGATCCGG 38

Db 16 AAGCTTGATCGCTGCAGGTGACTCTAGAGGATCCGG 53
 ||||||||||||||||||||||||||||||||||||||||

RESULT 6.

109132

LOCUS

DEFINITION

ACCESSION

VERSION

KEYWORDS

SOURCE

ORGANISM

REFERENCE

AUTHORS

TITLE

JOURNAL

FEATURES

source

BASE COUNT

ORIGIN

Query Match

Best Local Similarity

Matches

38; Conservative

0; Mismatches

0; Indels

0; Gaps

0;

QY 1 AAGCTTGATCGCTGCAGGTGACTCTAGAGGATCCGG 38

||||||||||||||||||||||||||||||||||||||

Db 18 AAGCTTGATCGCTGCAGGTGACTCTAGAGGATCCGG 55

RESULT 7

MTGLMS

LOCUS

DEFINITION

ACCESSION

VERSION

KEYWORDS

SOURCE

ORGANISM

REFERENCE

AUTHORS

TITLE

JOURNAL

REFERENCE

AUTHORS

TITLE

JOURNAL

FEATURES

source

1..2442

/organism="Mycobacterium tuberculosis"

/db_xref="taxon:1773"

/clone="genomic"

/clone_lib="cosmid"

1..2442

/gene="glms"

1..2442

/gene="glms"

/function="glucosamine fructose-6-phosphate

aminotransferase"

/codon_start=1

/transl_table=11

/protein_id="CAA04007.1"

/db_xref="GI:2275059"

gene

CDS

Query Match

Best Local Similarity

Matches

38; Conservative

0; Mismatches

0; Indels

0; Gaps

0;

QY 1 AAGCTTGATCGCTGCAGGTGACTCTAGAGGATCCGG 38

||||||||||||||||||||||||||||||||||||||

```
/db_xref="sptrembl:O33274"
/translation="MTMTIPSLHACRSTLEDPGHPSTLAVTWRMGPVGPRTTVPVTQS
RPGRVMTIPARRYSGPGWMDGVRVNGRPGVGGSHPPNRIEVGNORPDVHREERH
OTPTNEQVCLRLVRFALLARPDPGHVALLPODORTVPODELQVLGLVWGOVLA
FRKWPNSTIASVGRRCRCPDHAALPWSAVCGVIGVGRPAIVYVWMDALRRMEYRGY
DSSGIALVDGTLVRRRAGRLANLEEAEMPTSLGTTGLGHTRWATHGRPTDRN
ADHVRDAAGLAVVHNGIIEFVAVRRELETAGFEFASDTDEVAHHLVARAYRHGET
APPRGVSVALVRLRLEGHTLVFANADDPGTLVAAARRSTPLVLIGDNEMFVGSDDAA
FIEHTREAVELGQDAVVITADGVYRISDFDNDGLQAGRDFRPHIDWDLAAAEKGGY
EYFMLEKTAQPAOVAATLLGFGVGRIVLDEORLSQDELREIDKVPVACGTAYHSG
LKALIAEHTRLPVEVLELASEFRYROPVLDRLSRLVVAISQSGETADTLEAVRHAKQ
KAKYLAICENTNGSIIPRECDAYLVTRAGPEIGVASTKTFEQAIAANTLLGLALAQAAG
TKYDPEVEREYHELEAMPDLVARVIAATGPVLAELAHFAQSSTVFLGRHVGPVVALE
GALKLKLAYMHAGCFAGELKHGPIALIEDGLPVIIVMPSPKGSLTHAKLLSNIRE
IOTRGAVTIVIAEGDETVPYADHLEIIPAVSTLLQPLLSTIPLQVFAASVARARGY
DVDKPNLAKSVTVE"
BASE COUNT 415 a 800 c 815 g 412 t
ORIGIN
Query Match 0.9%; Score 38; DB 1; Length 2442;
Best Local Similarity 100.0%; Pred. No. 1.le-07;
Matches 38; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
Oy 1 AAGCTTGATCGCTGCAGGTCGACTCTAGAGGATCCGG 38
|||||
Db 18 AAGCTTGATCGCTGCAGGTCGACTCTAGAGGATCCGG 55
|||||

RESULT 8
SYNCRATBLB 4404 bp DNA linear SYN 16-FEB-1994
LOCUS
DEFINITION
Chloramphenicol acetyltransferase (CAT) and beta-lactamase genes,
complete cds.
ACCESSION M80483
VERSION M80483.1 GI:208101
KEYWORDS
CAT vector; beta-lactamase; chloramphenicol acetyltransferase;
thymidine kinase promoter; transient transfection vector.
SOURCE
Cloning vector DNA.
ORGANISM
artificial cloning vector
unidentified sequences; vectors.
REFERENCE
1 (bases 1 to 4404)
AUTHORS Boshart,M., Kluppel,M., Schmidt,A., Schutz,G. and Luckow,B.
TITLE Reporter constructs with low background activity utilizing the cat
gene
JOURNAL Gene 110 (1), 129-130 (1992)
MEDLINE 92184107
PUBMED 1544570
FEATURES
source
Location/Qualifiers
1..4404
/organism="unidentified cloning vector"
/db_xref="taxon:45196"
31..200
/size="200"
242..901
/size="659"
242..901
/size="659"
/size="659"
/codon_start=1
/transl_table=11
/product="chloramphenicol acetyltransferase"
/protein_id="AAA53138.1"
/db_xref="GI:208102"
/translation="MEKKITGYTTVDISQWHRKEHFEAFQSVQACTYNOTVOLDITAF
SLTKVKKHKEYPAFIHLARLMAHPEFRMAMKDGELVINDSVHPCTVTFHEQTETP
LKLKSEYHDDFROFLHIYSDVACYGENLAYEPKGFENMFVSANPWVSTFDLNV
AMNDFAPVTMGKYTQGDVLMPLAIQVHVAACDGFHVGRLMNELOQYCDDEWQGG
A"
991..1604
/misc_feature
/note="SV40 small T splice site"
1601..1841
/polyA_signal
/note="SV40 polyadenylation signal"
complement(3038..3898)
/gene="bla"

/db_xref="sptrembl:O33274"
/translation="MTMTIPSLHACRSTLEDPGHPSTLAVTWRMGPVGPRTTVPVTQS
RPGRVMTIPARRYSGPGWMDGVRVNGRPGVGGSHPPNRIEVGNORPDVHREERH
OTPTNEQVCLRLVRFALLARPDPGHVALLPODORTVPODELQVLGLVWGOVLA
FRKWPNSTIASVGRRCRCPDHAALPWSAVCGVIGVGRPAIVYVWMDALRRMEYRGY
DSSGIALVDGTLVRRRAGRLANLEEAEMPTSLGTTGLGHTRWATHGRPTDRN
ADHVRDAAGLAVVHNGIIEFVAVRRELETAGFEFASDTDEVAHHLVARAYRHGET
APPRGVSVALVRLRLEGHTLVFANADDPGTLVAAARRSTPLVLIGDNEMFVGSDDAA
FIEHTREAVELGQDAVVITADGVYRISDFDNDGLQAGRDFRPHIDWDLAAAEKGGY
EYFMLEKTAQPAOVAATLLGFGVGRIVLDEORLSQDELREIDKVPVACGTAYHSG
LKALIAEHTRLPVEVLELASEFRYROPVLDRLSRLVVAISQSGETADTLEAVRHAKQ
KAKYLAICENTNGSIIPRECDAYLVTRAGPEIGVASTKTFEQAIAANTLLGLALAQAAG
TKYDPEVEREYHELEAMPDLVARVIAATGPVLAELAHFAQSSTVFLGRHVGPVVALE
GALKLKLAYMHAGCFAGELKHGPIALIEDGLPVIIVMPSPKGSLTHAKLLSNIRE
IOTRGAVTIVIAEGDETVPYADHLEIIPAVSTLLQPLLSTIPLQVFAASVARARGY
DVDKPNLAKSVTVE"
BASE COUNT 415 a 800 c 815 g 412 t
ORIGIN
Query Match 0.9%; Score 38; DB 1; Length 2442;
Best Local Similarity 100.0%; Pred. No. 1.le-07;
Matches 38; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
Oy 1 AAGCTTGATCGCTGCAGGTCGACTCTAGAGGATCCGG 38
|||||
Db 18 AAGCTTGATCGCTGCAGGTCGACTCTAGAGGATCCGG 55
|||||

RESULT 9
BLCAT2DNA 4496 bp DNA linear SYN 26-FEB-1992
LOCUS
DEFINITION
Plasmid pBLCAT2 gene for beta-lactamase and CAT gene for
chloramphenicol acetyltransferase.
ACCESSION X64410
VERSION X64410.1 GI:58160
KEYWORDS
beta-lactamase; cat gene; chloramphenicol acetyltransferase;
thymidine kinase promoter.
SOURCE
synthetic construct.
artificial sequences.
1 (bases 1 to 4496)
REFERENCE
Luckow,B.H.R.
AUTHORS Direct Submission
TITLE Submitted (05-FEB-1992) B.H.R. Luckow, German Cancer Res Center, Im
Neuenheimer Feld 280, W-6900 Heidelberg, FRG
JOURNAL
REFERENCE
2 (bases 1 to 4496)
AUTHORS Luckow,B. and Schutz,G.
TITLE CAT constructions with multiple unique restriction sites for the
functional analysis of eukaryotic promoters and regulatory elements
Nucleic Acids Res. 15 (13), 5490 (1987)
JOURNAL 87260024
MEDLINE 3037497
PUBMED
COMMENT
CAT vector for transient transfection into eukaryotic cells.
Contains tk promoter for analysis of cis-acting elements in front
of a heterologous promoter.
FEATURES
source
Location/Qualifiers
1..4496
/organism="synthetic construct"
/db_xref="taxon:32630"
429..598
/promoter
/note="Herpes simplex virus tk promoter"
640..1299
/gene="CAT"
640..1299
/gene="CAT"
/codon_start=1
/transl_table=11
/product="chloramphenicol
acetyltransferase"
/protein_id="CAA45754.1"
/db_xref="GI:58161"
/translation="MEKKITGYTTVDISQWHRKEHFEAFQSVQACTYNOTVOLDITAF
LTKVKKHKEYPAFIHLARLMAHPEFRMAMKDGELVINDSVHPCTVTFHEQTETP
```


SSLSWEYHDDFRQFLHIYSODVACYGENLAYFPKGFIEENMFVFSANPWVSFTSLDLY
ANMNOFFAPVETMGKYYTQGDVKVLMPLAIQVHHAVCDFHVGRLMNELQYCDWEQGG
A"

misc_feature 1389..2002
/note="SV40 small T intron"
polyA_signal 1999..2239
/note="SV40 signals"
CDS complement(3436..4296)
/codon_start=1
/transl_table=11
/product="beta-lactamase"
/protein_id="CAA45753.1"
/db_xref="GI:58162"
/translation="MSIQHFRVALIPFFAFCFVFAHPETLVKVKDAEDOLGARVGY
IELDLSNGKILSEFRPEERFPMSTFKVLLCGAVLSRIDAGQEOQGRIRHYSONDLVE
YSPYTEKHLTDGTMVRELCSAAITMSDNTAANLLLTITIGGPKELTAFLLHNGDHTVRL
DRWPELNEAIPNDRDTPMVAATTLIKLLTGELTLLASRQQLIDMWEADKVGAPL
LRSLAPAGWFIADKSGAGERSGRIIAALGPDGPKPSRIIVYITTGSOATMDERNRQIA
EIGASLIKHW"

BASE COUNT 1188 a 1045 c 1070 g 1193 t

ORIGIN

Query Match 0.9%; Score 38; DB 12; Length 4496;

Best Local Similarity 100.0%; Pred. No. le-07;

Matches 38; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 AAGCTTGATCGCTGCAGTGCAGTCTAGAGGATCCGG 38

Db 399 AAGCTTGATCGCTGCAGTGCAGTCTAGAGGATCCGG 436

RESULT 10

AF027128

LOCUS

DEFINITION Eukaryotic luciferase expression vector ptkLUC+, complete sequence.

ACCESSION AF027128

VERSION AF027128.1

KEYWORDS

SOURCE

ORGANISM

REFERENCE

AUTHORS

TITLE

JOURNAL

MEDLINE

PUBMED

9298214

REFERENCE

AUTHORS

TITLE

JOURNAL

Submitted (26-SEP-1997)

Physiological Chemistry I/Biocenter,

University of Wuerzburg, Am Hubland, Wuerzburg 97074, Germany

Location/Qualifiers

1..4711

/organism="Expression vector ptkLUC+"

/db_xref="taxon:68901"

/plasmid="ptkLUC+"

1..36

/note="polylinker"

37..200

/note="tk-promoter"

238..1890

/gene="luc+"

238..1890

/gene="luc+"

/codon_start=1

/product="luciferase"

/protein_id="AAB83991.1"

/db_xref="GI:2598104"

/translation="MEDAKNIKGPAPFPYPLEDGTAGEQLHAKMKRYALVPCTIAFTD

AHVEDITYABYFEMSFLAEMRYGLNTNHRIVVCSNSLQFPMPLGALFIGVAV

APANDIYNRELLSMGISQPTVVVFSKGLQKLTINVKOKLPITIKIIMDSKTDYOG

FQSMYFTVSHLPFCFNEYDFVPESFDRDKTIALIMSSSGTGFLPKGVALPHRTACVR
FSHARDPIFGNOIIPDAILSVVPHHGFMTTGLYLICGFRVVLVYRFEELFLRS
LQDKIOSALLVPTLFSFAKSTLDKYNLHEIASGGAPLSKEVGEAAKPHLP
GIRGYGLTETTSAILITPEGDDAPCAVKVVPFEAKVYDLDTGKTLCGVNQGELCV
RGMIMSGYVNNPATNADIDKDWLHSGDIAYWDEDEHFFIVDLKSLIKYKGQVA
PAELISILQHPNIFDAGVAGLPDDAGELPAAVVVLEHGTMTKEIVDVVASQVTT
AKLRGGVVFVDEVPKGLTGKLDARKIREILIKAKKGKIAV"

misc_feature

misc_feature

gene

CDS

/note="splice and polyadenylation signals"

2146..2162

/note="polylinker"

complement(3345..4205)

/gene="bla"

complement(3345..4205)

/gene="bla"

/codon_start=1

/product="beta-lactamase"

/protein_id="AAB83992.1"

/db_xref="GI:2598105"

/translation="MSIQHFRVALIPFFAFCFVFAHPETLVKVKDAEDOLGARVGY
IELDLSNGKILSEFRPEERFPMSTFKVLLCGAVLSRIDAGQEOQGRIRHYSONDLVE
YSPYTEKHLTDGTMVRELCSAAITMSDNTAANLLLTITIGGPKELTAFLLHNGDHTVRL
DRWPELNEAIPNDRDTPMVAATTLIKLLTGELTLLASRQQLIDMWEADKVGAPL
LRSLAPAGWFIADKSGAGERSGRIIAALGPDGPKPSRIIVYITTGSOATMDERNRQIA
EIGASLIKHW"

misc_feature

BASE COUNT 1253 a 1102 c 1120 g 1236 t

ORIGIN

Query Match 0.9%; Score 38; DB 12; Length 4711;

Best Local Similarity 100.0%; Pred. No. le-07;

Matches 38; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 AAGCTTGATCGCTGCAGTGCAGTCTAGAGGATCCGG 38

Db 1 AAGCTTGATCGCTGCAGTGCAGTCTAGAGGATCCGG 38

RESULT 11

AF102233/c

LOCUS

DEFINITION Transposon delivery vector putkml, complete sequence.

ACCESSION AF102233

VERSION AF102233.1

KEYWORDS

SOURCE

ORGANISM

REFERENCE

AUTHORS

TITLE

JOURNAL

MEDLINE

PUBMED

2172217

REFERENCE

AUTHORS

TITLE

JOURNAL

Submitted (27-OCT-1998)

Department of Biochemistry, Stockholm

University, Svante Arrhenius vag. 10-12, Stockholm SE-10691, Sweden

Location/Qualifiers

1..7055

/organism="Transposon delivery vector putkml"

/db_xref="taxon:84477"

complement(311..1171)

/gene="bla"

complement(311..1171)

/gene="bla"

/codon_start=1

/transl_table=11

/product="beta lactamase"

Search completed: May 19, 2003, 07:44:01
Job time : 10419 secs

THIS PAGE BLANK (USPTO)

Result No.	Score	Query Match	Length	DB	ID	Description
1	4115	100.0	4115	21	AAA37430	Gluconobacter oxyd
2	38	0.9	1195	10	AAAS91189	Modified nitrilas
3	37	0.9	92	12	AAQ11697	Plasmid pWH332 fra
4	37	0.9	116	21	AAA10800	Nucleotide sequenc
5	37	0.9	586	12	AAQ11823	Region of pUC19-SA
6	37	0.9	586	15	AAQ78008	Bacteriophage T3 S
7	37	0.9	586	16	AAQ87073	Bacteriophage T3 S
8	37	0.9	586	17	AAT09985	S-adenosylmethioni
9	37	0.9	586	17	AAT09964	S-adenosylmethioni


```

XX PD 26-JAN-2000.
XX PF 17-JUL-1998; 98CN-0116333.
XX PR 17-JUL-1998; 98CN-0116333.
XX PA (HUAC-) HUACHEN BIOLOGICAL TECH RES INST SHANGHA.
XX PI Cao X, Zhao Z, Huang X;
XX PS WPI: 2000-351200/31.
XX PT Generation, cloning and sequence determination of a vector, and the
XX use thereof
XX PS Example 1; Fig 3; 16pp; Chinese.
XX CC This sequence represents the nucleotide sequence of pJL6. The invention
XX relates to methods for producing a vector containing the following
XX elements:
XX CC (i) a promoter for controlling exogenous gene transcription;
XX CC (ii) a reverse sequencing primer sequence;
XX CC (iii) an RBS (ribosome binding site) sequence;
XX CC (iv) a multiple cloning site directly downstream of the SD sequence;
XX CC (v) a positive sequencing primer sequence;
XX CC (vi) a terminator for controlling exogenous gene termination.
XX CC The multiple cloning site contains the enzymic incision site of a blunt
XX end type restriction endonuclease, and the interval of SD-AUG is 5-8
XX nucleotides. This method may be used to produce other vectors for
XX sequencing, cloning, and the expression of exogenous genes in the
XX molecular biology field.
XX SQ Sequence 116 BP; 32 A; 24 C; 28 G; 32 T; 0 other;

Query Match 0.9%; Score 37; DB 21; Length 116;
Best Local Similarity 100.0%; Pred. No. 4.2e-07;
Matches 37; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 AAGCTTCATGCCTGCAGTGCAGTCTAGAGGATCCG 37
    |||||
DB 91 AAGCTTCATGCCTGCAGTGCAGTCTAGAGGATCCG 55

RESULT 5
AAQ11823
ID AAQ11823 standard; DNA; 586 BP.
AC AAQ11823;
XX DT 24-SEP-1991 (first entry)
XX DE Region of pUC19-SAM-K vector.
XX KW S-adenosyl methionine hydrolase; ethylene biosynthesis; AdoMet;
XX KW homoserine; MTA; ss.
XX OS Bacteriophage T3.
XX FH Key Location/Qualifiers
XX FT CDS 66..524
XX FT /*tag= a
XX FT /label= AdoMetase
XX PN W09109112-A.
XX PD 27-JUN-1991.
XX PF 12-DEC-1990; 90WO-US07175.
XX PR 12-DEC-1989; 89US-0448095.
XX PA (AGRI-) AGRITOPIC INC.

```

```

XX PI Ferro AJ, Bestwick RK, Brown L;
XX DR WPI: 1991-208140/28.
XX DR P-PSDB; AAR12520.
XX PT Control of ethylene biosynthesis in plants - by introducing gene
XX which encodes S-adenosyl methionine hydrolase activity
XX PS Disclosure; Fig 6; 42pp; English.
XX CC The sequence is a fragment of vector pUC19-SAM-K for expression of
XX the AdoMetase gene. The gene is used to control ethylene biosyn-
XX thesis in plants. The enzyme hydrolyses S-adenosylmethionine to
XX homoserine and 5'-methylthioadenosine (MTA). The presence of the
XX gene and expression of the enzyme in plants lowers levels of
XX AdoMet, the sole precursor for ethylene biosynthesis, and its
XX reduced availability causes a corresponding decrease in ethylene
XX biosynthesis. The hydrolysis of AdoMet by the enzyme generates
XX MTA which is an inhibitor of 1-aminocyclopropane-1-carboxylic
XX acid (ACC)synthase, a principle enzyme in the biosynthesis of
XX ethylene. This results in fruit, vegetables and flowers which
XX have improved shelf life and preservation qualities.
XX CC See also AAQ11822.
XX SQ Sequence 586 BP; 144 A; 146 C; 157 G; 139 T; 0 other;

Query Match 0.9%; Score 37; DB 12; Length 586;
Best Local Similarity 100.0%; Pred. No. 3.9e-07;
Matches 37; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 AAGCTTCATGCCTGCAGTGCAGTCTAGAGGATCCG 37
    |||||
DB 24 AAGCTTCATGCCTGCAGTGCAGTCTAGAGGATCCG 60

RESULT 6
AAQ78008
ID AAQ78008 standard; DNA; 586 BP.
AC AAQ78008;
XX DT 05-JUL-1995 (first entry)
XX DE Bacteriophage T3 SAM-K modified S-adenosylmethionine hydrolase.
XX KW Tomato E8 promoter; transgenic fruit; ethylene production inhibition;
XX KW S-adenosylmethionine hydrolase; delayed fruit ripening; ds.
XX OS Bacteriophage T3.
XX FH Key Location/Qualifiers
XX FT CDS 66..524
XX FT /*tag= a
XX FT /label= a
XX PN W09424294-A.
XX PD 27-OCT-1994.
XX PF 08-APR-1994; 94WO-US03886.
XX PR 09-APR-1993; 93US-0046583.
XX PA (EPIT-) EPITOPE INC.
XX PI Bestwick RK, Ferro AJ;
XX DR WPI: 1994-341873/42.
XX DR P-PSDB; AAR74824.
XX PT Transgenic fruit-bearing plants, esp. tomato - contg. exogenous
XX gene under control of E8 promoter
XX

```

PS Claim 18; Fig 11; 59pp; English.

XX AA078007 the tomato E8 gene promoter, and AA078008 which encodes
 CC AAR74824 the bacteriophage T3 derived SAM-K modified
 CC S-adenosylmethionine hydrolase (AdoMetase), were used in the construction
 CC of a vector. The vector was used to produce transgenic tomato plants
 CC which expressed AdoMetase under the translational control of the E8
 CC promoter. AdoMetase is an inhibitor of ethylene production, and
 CC therefore an inhibitor of fruit ripening.

XX Sequence 586 BP; 144 A; 146 C; 157 G; 139 T; 0 other;

Query Match 0.9%; Score 37; DB 15; Length 586;

Best Local Similarity 100.0%; Pred. No. 3.9e-07;

Matches 37; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 AAGCTTGATGCTGCAGGTCGACTCTAGAGGATCCG 37
 |||||
 DB 24 AAGCTTGATGCTGCAGGTCGACTCTAGAGGATCCG 60

RESULT 7

AA087073

ID AA087073 standard; DNA; 586 BP.

XX AA087073;

DT 20-NOV-1995 (first entry)

DE Bacteriophage T3 S-adenosylmethionine-hydrolase gene.

KW Bacteriophage T3; phase T3; S-adenosylmethionine-hydrolase;

KW AdoMetase; SAM-hydrolase; enzyme; senescence; ethylene; hormone;

KW shelf-life; preservation; ripening; fruit; flower;

KW plasmid pUC19SAM-K; ss.

XX Bacteriophage T3.

XX Key Location/Qualifiers

FT cds 66..522

FT /*tag= a

XX US5416250-A.

XX 16-MAY-1995.

XX 12-DEC-1989; 89US-0448095.

XX 12-DEC-1989; 89US-0448095.

XX 12-DEC-1990; 90US-0613858.

XX 08-JUN-1994; 94US-0255833.

XX (AGRI-) AGRITOP INC.

XX Bestwick RK, Brown LR, Ferro AJ;

XX WPI; 1995-193491/25.

DR P-PSDB; AAR75212.

PT Reducing ethylene biosynthesis in plants - by introducing DNA
 encoding a S-adenosylmethionine-hydrolase enzyme into a plant host.

XX Disclosure; Fig. 6; 16pp; English.

XX This DNA encodes an S-adenosylmethionine-hydrolase enzyme which

CC hydrolyzes SAM to homoserine and 5'-methylthioadenosine. A

CC transgenic plant containing this sequence displays reduced ethylene

CC biosynthesis resulting in fruits, vegetables and flowers with

CC improved shelf life and preservation qualities.

XX Sequence 586 BP; 144 A; 146 C; 157 G; 139 T; 0 other;

Query Match 0.9%; Score 37; DB 16; Length 586;

Best Local Similarity 100.0%; Pred. No. 3.9e-07;

Matches 37; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 AAGCTTGATGCTGCAGGTCGACTCTAGAGGATCCG 37

|||||
 DB 24 AAGCTTGATGCTGCAGGTCGACTCTAGAGGATCCG 60

RESULT 8

AAT09985

ID AAT09985 standard; cDNA to mRNA; 586 BP.

XX AAT09985;

DT 11-AUG-1996 (first entry)

DE S-adenosylmethionine-hydrolase gene from pUC19-SAM-K.

XX S-adenosylmethionine-hydrolase; phase T3; chimeric gene;

KW tomato; E4 promoter; E8 promoter; fruit; transgenic plant;

KW ripening; ethylene; tissue-specific gene expression;

KW crop improvement; ds.

XX Bacteriophage T3 (ATCC 11303-B3).

XX Key Location/Qualifiers

FT misc_feature 55..83

FT /*tag= a

FT /note= "Region modified by linker to contain a
 eukaryotic translation initiation site"

FT CDS 66..524

FT /*tag= b

FT /product= S-adenosylmethionine-hydrolase

PN WO9535387-A1.

XX 28-DEC-1995.

XX 27-OCT-1994; 94WO-US12364.

XX 17-JUN-1994; 94US-0261677.

XX (EPIT-) EPITOPE INC.

XX Bestwick RK, Ferro AJ;

XX WPI; 1996-058424/06.

DR P-PSDB; AAR88611.

XX Transgenic plants with modified ripening phenotype - contg. a
 chimeric gene contg. a sequence which encodes a prod. which reduces
 ethylene biosynthesis

XX Example 1; Page 88-89; 139pp; English.

XX The sequence represents a modified S-adenosylmethionine-hydrolase
 (SAM) gene isolated from phase T3 and cloned in plasmid pUC19-SAM-K
 (plasmid PAG-111). A synthetic double-stranded oligonucleotide
 linker has been used to introduce a eukaryotic translation
 CC initiation site for expression in transgenic plants. The SAM gene,
 CC which is involved in ethylene biosynthesis and fruit ripening, may
 CC be inserted in a vector under the control of a plant ripening- or
 CC ethylene biosynthesis-inducible promoter (e.g. an E4 or E8
 CC promoter, as in AAT09973) and introduced into e.g. tomato to produce
 CC a transgenic plant bearing fruit with altered fruit ripening
 CC phenotype, resulting from an initial burst of ethylene production
 CC and a subsequent reduction in ethylene biosynthesis.

XX Sequence 586 BP; 144 A; 146 C; 157 G; 139 T; 0 other;

Query Match 0.9%; Score 37; DB 17; Length 586;

Best Local Similarity 100.0%; Pred. No. 3.9e-07;

Matches 37; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 AAGCTTCATGCCCTGCAGGTCGACTCTAGAGGATCCG 37
 |||||||||||||||||||||||||||||||||||||||||
 Db 24 AAGCTTCATGCCCTGCAGGTCGACTCTAGAGGATCCG 60

RESULT 9

AAT09964
 ID AAT09964 standard; cDNA to mRNA; 586 BP.

AC AAT09964;

DT 13-MAY-1996 (first entry)

XX S-adenosylmethionine hydrolase SAM-K cDNA.

XX S-adenosylmethionine hydrolase; AdoMet hydrolase; AdoMetase; SAM-K;
 KW S-adenosyl cleaving enzyme; SAMase; raspberry; Rubus idaeus;
 KW strawberry; Fragaria ananassa; transgenic plant; fruit rot;
 KW disease resistance; fungus resistance; virus resistance; ethylene;
 KW ripening; crop improvement; ds.

XX Synthetic.

XX Key Location/Qualifiers
 FH CDS 66..523
 FT /*tag= a

XX W09535388-A1.

XX 28-DEC-1995.

XX 16-JUN-1995; 95WO-US07753.

XX 03-FEB-1995; 95US-0384556.

XX 17-JUN-1994; 94US-0263900.

XX (EPIT-) EPIPOPE INC.

XX Bestwick RK, Ferro AJ, Mathews HV;

XX WPI: 1996-058425/06.

XX P-PSDB: AAR88601.

XX Transgenic red raspberry and strawberry plants - produce fruit with
 PT increased sugar content, increased fungal or viral resistance, or
 PT reduced ethylene biosynthesis

XX Disclosure; Page 42-43; 85pp; English.

XX The coding sequence of the phage T3 S-adenosylmethionine hydrolase
 CC gene was altered (AAR09964) so that the encoded enzyme, designated
 CC SAM-K (AAR88601), had an isoleucine to valine substn. at position
 CC 2. This highly conservative change increased the translation
 CC efficiency of the gene in plants. Transgenic raspberry and
 CC strawberry plants carrying the altered gene (see also AAR09967) were
 CC obtd. showing a high-efficiency transformation and regeneration system.
 CC They showed reduced levels of ethylene biosynthesis in leaf tissue
 CC and fruit.

XX Sequence 586 BP; 144 A; 146 C; 157 G; 139 T; 0 other;

Query Match 0.9%; Score 37; DB 17; Length 586;
 Best Local Similarity 100.0%; Pred. No. 3.9e-07;
 Matches 37; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 AAGCTTCATGCCCTGCAGGTCGACTCTAGAGGATCCG 37

Db 24 AAGCTTCATGCCCTGCAGGTCGACTCTAGAGGATCCG 60
 |||||||||||||||||||||||||||||||||||||||||

RESULT 10

AAT58564

ID AAT58564 standard; DNA; 586 BP.

AC AAT58564;

DT 17-SEP-1997 (first entry)

XX DNA sequence of AdoMetase gene with modified 5' end.

XX AdoMetase; bacteriophage T3; ethylene biosynthesis; transformation;
 KW S-adenosylmethionine hydrolase; hydrolysis; S-adenosylmethionine;
 KW AdoMet; homoserine; 5'-methylthioadenosine; MTA; ornamental plant;
 KW 1-aminocyclopropane-1-carboxylic acid; ACC; inhibitor; ACC synthase;
 KW transgenic plant; fruit; vegetable; flower; shelf life; preservation;
 KW food; ss.

XX Bacteriophage T3.

OS Synthetic.

XX Key Location/Qualifiers

FH CDS 66..524

FT /*tag= a

FT /product= AdoMetase

XX US5589623-A.

XX 31-DEC-1996.

XX 12-DEC-1989; 89US-0448095.

XX 12-DEC-1990; 90US-0613858.

XX 12-DEC-1989; 89US-0448095.

XX 08-JUN-1994; 94US-0255833.

XX 20-DEC-1994; 94US-0360974.

XX (AGRI-) AGRITPOPE INC.

XX Bestwick RK, Brown LR, Ferro AJ;

XX WPI: 1997-076910/07.

XX P-PSDB: AAW11594.

XX Control of ethylene biosynthesis in plants using vector encoding
 PT S-adenosyl:methionine hydrolase - used to produce fruit, flowers
 PT etc. of increased shelf life and preservation quality

XX Example 3; Fig 6; 16pp; English.

XX This sequence represents the AdoMetase gene derived from bacteriophage
 CC T3. This sequence may be used in the method of the invention to control
 CC ethylene biosynthesis in plants. This is done by transforming plant
 CC cells with a vector which contains a DNA sequence (I) allowing genetic
 CC selection in plant cells and a DNA sequence (II) encoding S-adenosyl-
 CC methionine hydrolase (AdoMetase) which hydrolyses S-adenosylmethionine
 CC (AdoMet) to homoserine and 5'-methylthioadenosine (MTA). Both (I) and
 CC (II) are flanked by regulatory elements that allow expression in plant
 CC cells. AdoMetase reduces levels of AdoMet, the essential precursor of
 CC ethylene (via 1-aminocyclopropane-1-carboxylic acid (ACC)) and MTA is
 CC an inhibitor of ACC synthase, the major enzyme involved in ethylene
 CC synthesis. Transgenic plants containing this construct produce fruits,
 CC vegetables and flowers of improved shelf life and preservation qualities.
 CC A wide range of food and ornamental plants can be modified. The
 CC transgenic plants are able to control ethylene production under
 CC restricted conditions. This sequence has been modified around the
 CC initiation codon to produce an optimised kozak sequence.

XX Sequence 586 BP; 144 A; 146 C; 157 G; 139 T; 0 other;

Query Match 0.9%; Score 37; DB 18; Length 586;
 Best Local Similarity 100.0%; Pred. No. 3.9e-07;
 Matches 37; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 AAGCTTCATGCCCTGCAGGTCGACTCTAGAGGATCCG 37

|||||||||||||||||||||||||||||||||||||||||

Db 24 AAGCTTGACGCTGCAGGTGCACTCTAGAGGATCCG 60

RESULT 11

AAAX00468
ID AAX00468 standard; cDNA to mRNA; 586 BP.

XX AC AAX00468;

XX DT 31-WAR-1999 (first entry)

XX DE S-adenosylmethionine hydrolase modified gene.

XX KW Ethylene; plant hormone; regulator; metabolism; senescence; E4; E8;
KW S-adenosylmethionine hydrolase; SAMase; promoter; fruit ripening;
KW storage; ds.

XX OS Bacteriophage t3.

XX OS Synthetic.

XX FH Key Location/Qualifiers
FT CDS 66..524
FT /*tag= a

XX PN US5859330-A.

XX PD 12-JAN-1999.

XX PF 27-OCT-1994; 94US-0331355.

XX PR 27-OCT-1994; 94US-0331355.

XX PR 12-DEC-1989; 89US-0448095.

XX PR 12-DEC-1990; 90US-0613858.

XX PR 09-APR-1993; 93US-0046583.

XX PR 08-JUN-1994; 94US-0255833.

XX PR 17-JUN-1994; 94US-0261677.

XX PA (EPIT-) EPITOPE INC.

XX PI Bestwick RK, Ferro AJ;

XX WPI: 1999-120030/10.

XX DR P-PSDB; AAW30604.

XX Transgenic plants containing S-adenosylmethionine hydrolase gene -
PT useful for delaying fruit ripening and lengthening storage periods

XX PS Example 1; Fig 11; 72pp; English.

XX The present invention describes a transgenic fruit-bearing plant
CC containing a DNA sequence encoding S-adenosylmethionine hydrolase
CC (SAMase) linked to a promoter selected from a group consisting of: (a)
CC a tomato E4 promoter; (b) a tomato E8 promoter; and (c) an avocado
CC cellulase promoter; (d) a tomato E4 promoter/SAMase construct.
CC Expression from the promoter is induced during fruit ripening,
CC optionally by ethylene synthesised in the fruit. The SAMase transgene
CC produces fruit with reduced ethylene synthesis, delaying fruit ripening
CC and lengthening storage periods. The present sequence represents a
CC modified SAMase gene from the present invention.

XX SQ Sequence 586 BP; 144 A; 146 C; 157 G; 139 T; 0 other;

Query Match 0.9%; Score 37; DB 20; Length 586;

Best Local Similarity 100.0%; Pred. No. 3.9e-07;

Matches 37; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 AAGCTTGACGCTGCAGGTGCACTCTAGAGGATCCG 37

XXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXX

Db 24 AAGCTTGACGCTGCAGGTGCACTCTAGAGGATCCG 60

RESULT 12

AAA27831/c

ID AAA27831 standard; DNA; 3796 BP.
XX AC AAA27831;
XX DT 12-SEP-2000 (first entry)
XX DE Vector plasmid pCMV-MC1.
XX KW North American PRRS virus; Nidovirales virus; pig; swine; vaccine;
XX KW pCMV-S-P129; ss.
XX OS Chimeric - Porcine reproductive and respiratory syndrome virus.
XX OS Chimeric - Human cytomegalovirus.
XX PN EP1018557-A2.
XX PD 12-JUL-2000.
XX PF 25-NOV-1999; 99EP-0309409.
XX PR 22-DEC-1998; 98US-0113345.
XX PA (PFIZ) PFIZER PROD INC.
XX PI Calvert JG, Welch SW, Sheppard MG;
XX WPI: 2000-444364/39.
XX New polynucleotide encoding an infectious RNA molecule of a North
PT American porcine reproductive and respiratory syndrome virus for use as
PT a vaccine in protecting swine and other animals from infection by a
PT pathogen -
XX Example 4; Page 43-44; 53pp; English.
XX The present sequence is that of eukaryotic expression vector
CC plasmid pCMV-MC1. The plasmid is derived from pCMVbeta by
CC replacing the LacZ coding sequence with a linker containing
CC multiple restriction sites. Modification of the human
CC cytomegalovirus (CMV) immediate early promoter was accomplished by
CC substituting the sequence between SacI and the 2nd NotI site of
CC pCMV-MC1 with a synthetic linker. North American porcine
CC reproductive and respiratory syndrome (PRRS) virus P129A cDNA (see
CC AAA27809) was cloned into the modified vector to create pCMV-S-P129
CC (ATCC 203489), which was used to demonstrate cellular expression of
CC PRRS virus by direct transfection of cDNA into cells. The
CC invention relates to polynucleotide molecules, plasmids, viral
CC vectors and transfected host cells that comprise North American
CC PRRS DNA. It also relates to polynucleotide molecules, viral
CC vectors and transfected host cells encoding a genetically modified
CC North American PRRS virus that is disabled in its ability to cause
CC PRRS, or which encodes 1 or more heterologous antigenic epitopes,
CC for use as a vaccine.
XX SQ Sequence 3796 BP; 953 A; 934 C; 956 G; 953 T; 0 other;
Query Match 0.9%; Score 37; DB 21; Length 3796;
Best Local Similarity 100.0%; Pred. No. 3.5e-07;
Matches 37; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 AAGCTTGACGCTGCAGGTGCACTCTAGAGGATCCG 37
XXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXX
Db 1167 AAGCTTGACGCTGCAGGTGCACTCTAGAGGATCCG 1131
XXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXX
RESULT 13
AAT15287
ID AAT15287 standard; cDNA; 4910 BP.
XX AC AAT15287;
XX DT 27-OCT-1996 (first entry)
XX

DE prEP78/52 hybrid adeno virus-adeno-associated virus cDNA.
 XX
 KW Adeno virus; adeno-associated virus; hybrid virus vector;
 KW reporter gene; transgene; somatic gene therapy; gene transfer;
 KW plasmid prEP78/52; familial hypercholesterolemia; cystic fibrosis;
 KW ds.
 XX
 OS Adeno virus; Adeno-associated virus.
 XX
 PN W09613598-A2.
 XX
 XX
 PD 09-MAY-1996.
 XX
 XX 27-OCT-1995; 95WO-US14018.
 PF
 XX 28-OCT-1994; 94US-0331384.
 PR
 XX (UYPE-) UNIV PENNSYLVANIA.
 PA
 XX Fisher KJ, Kelley WM, Wilson JM;
 PI
 XX WPI; 1996-239504/24.
 DR
 XX Hybrid adenovirus-adeno-associated virus for gene therapy -
 PT comprises adenovirus portion, 5' and 3' ITR sequences from AAV and a
 PT selected transgene, e.g. CFTR or LDL
 FT
 XX
 PS Disclosure; Page 57-61; 91pp; English.
 XX
 CC This cDNA encodes recombinant plasmid prEP78/52, a trans-acting
 CC plasmid containing the AAV sequences that encode rep 78 kD and
 CC 52 kD proteins under the control of the AAV P5 promoter. The
 CC plasmid also contains an SV40 polyadenylation signal. This cDNA
 CC is used in the construction of a hybrid AAV-AV virus vector. The
 CC hybrid virus may be used in the delivery and stable integration
 CC of a selected gene into the chromosome of a target cell, i.e. gene
 CC therapy of cystic fibrosis and familial hypercholesterolaemia.
 XX
 SQ Sequence 4910 BP; 1161 A; 1294 C; 1215 G; 1240 T; 0 other;
 XX
 Query Match 0.9%; Score 37; DB 17; Length 4910;
 Best Local Similarity 100.0%; Pred. No. 3.4e-07;
 Matches 37; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 OY 1 AAGCTTCATGCGCTCGAGGTCGACTCTAGAGGATCCG 37
 |||||
 DB 399 AAGCTTCATGCGCTCGAGGTCGACTCTAGAGGATCCG 435
 RESULT 14
 AAV02043/c
 ID AAV02043 standard; DNA; 6295 BP.
 XX
 XX AAV02043;
 XX
 DT 08-JUN-1998 (first entry)
 DE
 XX Plasmid pWRG3196 encoding murine interleukin-12.
 KW
 KW Interleukin-12; IL-12; cytokine; growth factor; mouse1
 KW plasmid pWRG3196; cancer; tumour; metastasis; gene therapy; ds;
 KW cyclic; circular.
 XX
 OS Chimeric - Mus musculus.
 OS Chimeric - Cytomegalovirus.
 OS Chimeric - Bos taurus.
 OS Chimeric - Rhesus macaque polyoma virus
 OS Chimeric - encephalomyocarditis virus.
 XX
 XX Key Location/Qualifiers
 FH 955..1675
 FT CDS /*tag= a
 FT /product= p35 subunit

FT intron /note= "contains an intron"
 FT 1261..1333
 FT /*tag= b
 FT CDS 2377..3384
 FT /*tag= c
 FT /product= p40 subunit
 XX
 PN W09746263-A1.
 XX
 XX 11-DEC-1997.
 PD
 XX 04-JUN-1997; 97WO-US09591.
 PF
 XX 05-JUN-1996; 96US-0659206.
 PR
 XX (AURA-) AURAGEN INC.
 PA
 XX Rakhmillevich AL, Yang N;
 PI
 XX WPI; 1998-041898/04.
 DR
 XX P-PSDB; AAW44004-05.
 DR
 XX Interleukin-12 gene therapy of tumours - comprises delivering
 PT construct containing promoter and sequences encoding interleukin-12
 PT p35 and p40 subunit(s) to target cells in vivo
 FT
 XX Claim 8; Page 33-38; 50pp; English.
 PS
 XX Plasmid pWRG3196 is a bicistronic plasmid encoding both subunits,
 CC i.e. p35 (see AAW44004) and p40 (see AAW44005), of murine interleukin-12
 CC (IL-12). The p35 and p40 subunits were cloned from a mouse spleen
 CC cDNA library. The vector contains a single cytomegalovirus promoter,
 CC an SV40 splicing donor/splicing acceptor, and bovine growth hormone
 CC polyA signal. Between the p35 and p40 genes is an internal
 CC ribosome entry site element (IRES) cloned from encephalomyocarditis
 CC virus. The backbone of the plasmid is from pUC19. Plasmid
 CC pWRG3139 induces about half the expression of IL-12 as the tandem
 CC vector pWRG3169 (see AAV02042) in vivo and in vitro. A novel method
 CC of treating tumours in a mammal involves delivering copies of an
 CC expressible foreign genetic construct, especially pWRG3169 or
 CC pWRG3196, comprising a promoter operative in the mammalian
 CC epidermal cells and DNA sequences encoding p35 and p40 subunits of
 CC IL-12 to target cells in vivo. Delivery of the construct allows
 CC IL-12 expression for treatment of solid, metastatic or disseminated
 CC tumours, and regression of established tumours. The treatment is
 CC effective even when the genetic construct is delivered to a site
 CC distant from the tumour.
 XX
 SQ Sequence 6295 BP; 1610 A; 1610 C; 1568 G; 1507 T; 0 other;
 XX
 Query Match 0.9%; Score 37; DB 19; Length 6295;
 Best Local Similarity 100.0%; Pred. No. 3.4e-07;
 Matches 37; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 OY 1 AAGCTTCATGCGCTCGAGGTCGACTCTAGAGGATCCG 37
 |||||
 DB 3666 AAGCTTCATGCGCTCGAGGTCGACTCTAGAGGATCCG 3630
 RESULT 15
 AAD10238/c
 ID AAD10238 standard; DNA; 7164 BP.
 XX
 XX AAD10238;
 XX
 DT 24-SEP-2001 (first entry)
 DE
 XX Commercial plasmid vector pCMVbeta.
 XX
 KW Plasmid; cyclic; circular; replicon; exogenous gene; marker gene;
 KW transcription termination; immunostimulatory sequence; ISS; antiviral;
 KW non-essential nucleotide; molecular biology application; gene therapy;
 KW DNA vaccine; cloning; gene expression; in vitro protein production;

KW cytostatic; pCMVbeta; cytomegalovirus promoter; lacZ gene;
 KW ampicillin gene; simian virus 40; SV40 intron; ds.
 XX
 OS Chimeric - Cytomegalovirus.
 OS Chimeric - Rhesus macaque polyoma virus.
 OS Chimeric - Unidentified.
 XX
 FH
 FT Location/Qualifiers
 FT misc_feature 1..26
 FT /tag= a
 FT /note= "Corresponds to the non-essential nucleotide
 FT sequence that have been removed in the novel DNA
 FT plasmid vector."
 FT 4513..4890
 FT misc_feature
 FT /tag= b
 FT /note= "pUC18 lacZ promoter that corresponds to the
 FT non-essential nucleotide sequence that have been removed
 FT in the novel DNA plasmid vector."
 FT 7007..7164
 FT misc_feature
 FT /tag= c
 FT /note= "5' non-coding region of lacZ gene from pUC18 that
 FT corresponds to the non-essential nucleotide sequence
 FT that have been removed in the novel DNA plasmid vector."
 XX
 PN WO200151626-A2.
 XX
 XX
 PD 19-JUL-2001.
 XX
 XX
 PF 09-JAN-2001; 2001WO-US01255.
 XX
 XX
 PR 10-JAN-2000; 2000US-0480879.
 XX
 PA (ELIM-) ELIM BIOPHARMACEUTICALS INC.
 XX
 PI Lu X, Sun L, Zhang Y;
 XX
 DR WPI; 2001-451855/48.
 XX
 XX New plasmid DNA vectors, useful for most molecular biology
 PT applications, e.g. gene therapy, DNA vaccines, cloning and expression
 PT of genes, and in the in vitro production of polypeptides and/or
 PT proteins -
 XX
 PS Example 2; Page 42-44; 50pp; English.
 XX
 CC The present invention relates to plasmid DNA vectors comprising
 CC essentially of a replicon and at least one other component selected from
 CC promoter, intron, exogenous gene, transcription termination sequence,
 CC selectable marker gene, detectable marker gene and an immunostimulatory
 CC sequence (ISS), where the non-essential nucleotide sequences have been
 CC substantially removed from these vectors. The plasmid DNA vectors are
 CC useful in most molecular biology applications, e.g. gene therapy, DNA
 CC vaccines, cloning and expression of genes, and in the in vitro production
 CC of polypeptides and/or proteins. The present sequence is a commercial
 CC plasmid DNA vector pCMVbeta which comprises Cytomegalovirus promoter,
 CC Simian Virus 40 (SV40) intron, the lacZ gene, pUC origin of replication
 CC and the ampicillin gene.
 XX
 SQ Sequence 7164 BP; 1700 A; 1840 C; 1897 G; 1727 T; 0 other;
 Query Match 0.9%; Score 37; DB 22; Length 7164;
 Best Local Similarity 100.0%; Pred. No. 3.3e-07;
 Matches 37; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 OY 1 AAGCTTCATCCCTGCAGGTCGACTCTAGAGGATCCG 37
 DB 4535 AAGCTTCATCCCTGCAGGTCGACTCTAGAGGATCCG 4499
 Search completed: May 19, 2003, 08:05:39
 Job time : 872 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: May 18, 2003, 22:35:30 ; Search time 178 Seconds
(without alignments)
7089.746 Million cell updates/sec

Title: US-09-926-163B-1

Perfect score: 4115

Sequence: 1 aagcttgatgcctgcaggt.....tccggcgccgctgaagctt 4115

Scoring table: OLIGO_NUC

Gapop 60.0 , Gapext 60.0

Searched: 441362 seqs, 153338381 residues

Word size : 0

Total number of hits satisfying chosen parameters: 882724

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Database : Issued_Patents_NA.*
1: /cgn2_6/ptodata/1/ina/5A_COMB.seq:*
2: /cgn2_6/ptodata/1/ina/5B_COMB.seq:*
3: /cgn2_6/ptodata/1/ina/6A_COMB.seq:*
4: /cgn2_6/ptodata/1/ina/6B_COMB.seq:*
5: /cgn2_6/ptodata/1/ina/PCTUS_COMB.seq:*
6: /cgn2_6/ptodata/1/ina/backfiles1.seq:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	37	0.9	586	1 US-08-046-583-10	Sequence 10, Appl
2	37	0.9	586	1 US-08-384-556A-1	Sequence 1, Appl
3	37	0.9	586	2 US-08-331-355A-22	Sequence 22, Appl
4	37	0.9	586	5 PCT-US94-12364-22	Sequence 22, Appl
5	37	0.9	586	5 PCT-US95-07753-1	Sequence 1, Appl
6	37	0.9	4910	2 US-08-331-384-2	Sequence 2, Appl
7	37	0.9	4910	2 US-08-836-087-2	Sequence 2, Appl
8	37	0.9	4910	4 US-09-246-320-2	Sequence 2, Appl
9	37	0.9	4910	4 US-09-546-738-2	Sequence 2, Appl
10	37	0.9	6295	2 US-08-659-206A-4	Sequence 4, Appl
11	37	0.9	7897	3 US-08-836-022A-1	Sequence 1, Appl
12	37	0.9	7897	4 US-09-427-048A-1	Sequence 1, Appl
13	37	0.9	8540	4 US-08-487-283A-4	Sequence 4, Appl
14	37	0.9	8540	5 PCT-US96-05811A-12	Sequence 12, Appl
15	37	0.9	8932	2 US-08-252-493C-8	Sequence 8, Appl
16	37	0.9	8932	3 US-09-276-197-8	Sequence 8, Appl
17	37	0.9	8932	3 US-08-836-022A-3	Sequence 3, Appl
18	37	0.9	9972	4 US-09-427-048A-3	Sequence 3, Appl
19	37	0.9	19307	3 US-08-836-022A-10	Sequence 10, Appl
20	37	0.9	19307	4 US-09-427-048A-10	Sequence 10, Appl
21	36	0.9	40	2 US-08-660-295A-4	Sequence 4, Appl
22	36	0.9	50	2 US-08-693-302-3	Sequence 3, Appl
23	36	0.9	50	4 US-09-099-466-3	Sequence 3, Appl
24	36	0.9	57	1 US-08-089-910-19	Sequence 19, Appl
25	36	0.9	57	1 US-08-089-910-22	Sequence 22, Appl
26	36	0.9	57	1 US-08-400-864-4	Sequence 4, Appl
27	36	0.9	57	4 US-09-116-492A-33	Sequence 33, Appl

Sequence 1, Appl
Patent No. 5166321
Sequence 66, Appl
Sequence 8, Appl
Sequence 5, Appl
Sequence 8, Appl
Patent No. 5304637
Patent No. 5166321
Sequence 73, Appl
Sequence 73, Appl
Sequence 1, Appl
Sequence 3, Appl
Sequence 1, Appl
Patent No. 5231168
Sequence 6, Appl
Sequence 7, Appl
Sequence 11, Appl

28 36 0.9 59 4 US-09-233-507-1
29 36 0.9 59 6 5166321-2
30 36 0.9 60 1 US-08-261-206A-66
31 36 0.9 60 4 US-09-586-935-8
32 36 0.9 60 4 US-08-470-389-5
33 36 0.9 60 4 US-09-214-146-8
34 36 0.9 60 6 5304637-5
35 36 0.9 63 6 5166321-4
36 36 0.9 64 2 US-08-484-956-73
37 36 0.9 64 2 US-08-757-653-73
38 36 0.9 64 4 US-08-520-946-73
39 36 0.9 66 1 US-07-729-460A-1
40 36 0.9 66 1 US-08-422-613-3
41 36 0.9 66 4 US-07-780-717C-1
42 36 0.9 66 6 5231168-15
43 36 0.9 72 1 US-08-373-134D-6
44 36 0.9 72 1 US-08-373-134D-7
45 36 0.9 72 2 US-08-585-585A-11

ALIGNMENTS

RESULT 1
US-08-046-583-10
; Sequence 10, Application US/08046583
; Patent No. 5723746
; GENERAL INFORMATION:
; APPLICANT: Bestwick, Richard K
; APPLICANT: Ferro, Adolph J
; TITLE OF INVENTION: Reduced Ethylene Synthesis and Delayed
; TITLE OF INVENTION: Fruit Ripening in Transgenic Tomatoes Expressing
; TITLE OF INVENTION: S-Adenosylmethionine Hydrolase
; NUMBER OF SEQUENCES: 12
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Denlinger & Associates
; STREET: 350 Cambridge Avenue, Suite 250
; CITY: Palo Alto
; STATE: CA
; COUNTRY: USA
; ZIP: 94306
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/046,583
; FILING DATE: 09-APR-1993
; CLASSIFICATION: 800
; ATTORNEY/AGENT INFORMATION:
; NAME: Fabian, Gary R.
; REGISTRATION NUMBER: 33,875
; REFERENCE/DOCKET NUMBER: 4257-0006
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (415) 324-0880
; TELEFAX: (415) 324-0960
; INFORMATION FOR SEQ ID NO: 10:
; SEQUENCE CHARACTERISTICS:
; .LENGTH: 586 base pairs
; .TYPE: nucleic acid
; .STRANDEDNESS: double
; .TOPOLOGY: linear
; .MOLECULE TYPE: cDNA to mRNA
; .HYPOTHETICAL: NO
; .ANTI-SENSE: NO
; .ORIGINAL SOURCE:
; .INDIVIDUAL ISOLATE: Fig. 11, pUC19-SAM-K
; .FEATURE:
; .NAME/KEY: CDS
; .LOCATION: 66..521
US-08-046-583-10

Query Match 0.9%; Score 37; DB 1; Length 586;
Best Local Similarity 100.0%; Pred. No. 2.6e-07;
Matches 37; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 AAGCTTCATGCTGCAGTGCAGTCTAGAGGATCCG 37
|||||
Db 24 AAGCTTCATGCTGCAGTGCAGTCTAGAGGATCCG 60

RESULT 2

US-08-384-556A-1
; Sequence 1, Application US/08384556A
; Patent No. 5750870
; GENERAL INFORMATION:
; APPLICANT: Mathews, Helena V
; APPLICANT: Bestwick, Richard K
; APPLICANT: Ferro, Adolph J
; TITLE OF INVENTION: Plant Genetic Transformation
; TITLE OF INVENTION: Methods and Transgenic Plants
; NUMBER OF SEQUENCES: 11
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Dehlinger & Associates
; STREET: 350 Cambridge Avenue, Suite 250
; CITY: Palo Alto
; STATE: CA
; COUNTRY: USA
; ZIP: 94306
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/384,556A
; FILING DATE:
; CLASSIFICATION: 800
; PRIOR APPLICATION DATA:
; PRIOR APPLICATION NUMBER: US 08/263,900
; FILING DATE: 17-JUN-1994
; APPLICATION NUMBER: US 08/263,900
; FILING DATE: 17-JUN-1994
; ATTORNEY/AGENT INFORMATION:
; NAME: Fabian, Gary R.
; REGISTRATION NUMBER: 33,875
; REFERENCE/DOCKET NUMBER: 4257-0010.30
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (415) 324-0880
; TELEFAX: (415) 324-0960
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 586 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; MOLECULE TYPE: cDNA to mRNA
; HYPOTHETICAL: NO
; ANTI-SENSE: NO
; ORIGINAL SOURCE:
; INDIVIDUAL ISOLATE: SAM-K
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 66..521
; US-08-384-556A-1

Query Match 0.9%; Score 37; DB 1; Length 586;
Best Local Similarity 100.0%; Pred. No. 2.6e-07;
Matches 37; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 AAGCTTCATGCTGCAGTGCAGTCTAGAGGATCCG 37
|||||
Db 24 AAGCTTCATGCTGCAGTGCAGTCTAGAGGATCCG 60

RESULT 3

US-08-331-355A-22

; Sequence 22, Application US/08331355A
; Patent No. 5859330
; GENERAL INFORMATION:
; APPLICANT: Bestwick, Richard K
; APPLICANT: Ferro, Adolph J
; TITLE OF INVENTION: Regulated Expression of Heterologous
; TITLE OF INVENTION: Genes in Plants and Transgenic Fruit
; TITLE OF INVENTION: with a Modified Ripening Phenotype
; NUMBER OF SEQUENCES: 27
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Dehlinger & Associates
; STREET: 350 Cambridge Avenue, Suite 250
; CITY: Palo Alto
; STATE: CA
; COUNTRY: USA
; ZIP: 94306
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/331,355A
; FILING DATE: 27-OCT-1994
; CLASSIFICATION: 800
; PRIOR APPLICATION DATA:
; PRIOR APPLICATION NUMBER: US 08/261,677
; FILING DATE: 17-JUN-1994
; APPLICATION NUMBER: US 08/046,583
; FILING DATE: 09-APR-1993
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/255,833
; FILING DATE: 08-JUN-1994
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/613,858
; FILING DATE: 12-DEC-1990
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 448,095
; FILING DATE: 12-DEC-1989
; ATTORNEY/AGENT INFORMATION:
; NAME: Evans, Susan T.
; REGISTRATION NUMBER: 38,443
; REFERENCE/DOCKET NUMBER: 4257-0011.30
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (415) 324-0880
; TELEFAX: (415) 324-0960
; INFORMATION FOR SEQ ID NO: 22:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 586 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; MOLECULE TYPE: cDNA to mRNA
; HYPOTHETICAL: NO
; ANTI-SENSE: NO
; ORIGINAL SOURCE:
; INDIVIDUAL ISOLATE: Fig. 11, pUC19-SAM-K
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 66..521
; US-08-331-355A-22

Query Match 0.9%; Score 37; DB 2; Length 586;
Best Local Similarity 100.0%; Pred. No. 2.6e-07;
Matches 37; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 AAGCTTCATGCTGCAGTGCAGTCTAGAGGATCCG 37
|||||
Db 24 AAGCTTCATGCTGCAGTGCAGTCTAGAGGATCCG 60

RESULT 4

PCT-US94-12364-22
; Sequence 22, Application PC/TUS9412364
; GENERAL INFORMATION:
; APPLICANT:
; TITLE OF INVENTION: Regulated Expression of Heterologous
; TITLE OF INVENTION: Genes in Plants and Transgenic Fruit
; TITLE OF INVENTION: with a Modified Ripening Phenotype
; NUMBER OF SEQUENCES: 26
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Dehlinger & Associates
; STREET: 350 Cambridge Avenue, Suite 250
; CITY: Palo Alto
; STATE: CA
; COUNTRY: USA
; ZIP: 94306
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: PCT/US94/12364
; FILING DATE:
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/261,677
; FILING DATE: 17-JUN-1994
; ATTORNEY/AGENT INFORMATION:
; NAME: Fabian, Gary R.
; REGISTRATION NUMBER: 33,875
; REFERENCE/DOCKET NUMBER: 4257-0011.41
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (415) 324-0880
; TELEFAX: (415) 324-0960
; INFORMATION FOR SEQ ID NO: 22:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 586 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; MOLECULE TYPE: cdna to mRNA
; HYPOTHETICAL: NO
; ANTI-SENSE: NO
; ORIGINAL SOURCE:
; INDIVIDUAL ISOLATE: Fig. 11, pUC19-SAM-K
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 66..521
; PCT-US94-12364-22

Query Match 0.9%; Score 37; DB 5; Length 586;
Best Local Similarity 100.0%; Pred. No. 2.6e-07;
Matches 37; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 AAGCTTGATCGCTGCAGTGCAGTCTAGAGGATCCG 37
|||||
Db 24 AAGCTTGATCGCTGCAGTGCAGTCTAGAGGATCCG 60

RESULT 5
PCT-US95-07753-1
; Sequence 1, Application PC/TUS9507753
; GENERAL INFORMATION:
; APPLICANT:
; TITLE OF INVENTION: Plant Genetic Transformation
; TITLE OF INVENTION: Methods and Transgenic Plants
; NUMBER OF SEQUENCES: 11
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Dehlinger & Associates
; STREET: 350 Cambridge Avenue, Suite 250
; CITY: Palo Alto
; STATE: CA
; COUNTRY: USA

ZIP: 94306
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: PCT/US95/07753
; FILING DATE:
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/384,556
; FILING DATE: 03-FEB-1995
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/263,900
; FILING DATE: 17-JUN-1994
; ATTORNEY/AGENT INFORMATION:
; NAME: Fabian, Gary R.
; REGISTRATION NUMBER: 33,875
; REFERENCE/DOCKET NUMBER: 4257-0010.41
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (415) 324-0880
; TELEFAX: (415) 324-0960
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 586 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; MOLECULE TYPE: cdna to mRNA
; HYPOTHETICAL: NO
; ANTI-SENSE: NO
; ORIGINAL SOURCE:
; INDIVIDUAL ISOLATE: SAM-K
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 66..521
; PCT-US95-07753-1

Query Match 0.9%; Score 37; DB 5; Length 586;
Best Local Similarity 100.0%; Pred. No. 2.6e-07;
Matches 37; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 AAGCTTGATCGCTGCAGTGCAGTCTAGAGGATCCG 37
|||||
Db 24 AAGCTTGATCGCTGCAGTGCAGTCTAGAGGATCCG 60

RESULT 6
US-08-331-384-2
; Sequence 2, Application US/08331384
; Patent No. 5856152
; GENERAL INFORMATION:
; APPLICANT: Wilson, James M.
; APPLICANT: Kelley, William M.
; APPLICANT: Fisher, Krishna J.
; TITLE OF INVENTION: Hybrid Adenovirus-AAV Vector and Methods
; TITLE OF INVENTION: of Use Thereof
; NUMBER OF SEQUENCES: 2
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Howson and Howson
; STREET: Spring House Corporate Cntr, P.O. Box 457
; CITY: Spring House
; STATE: Pennsylvania
; COUNTRY: USA
; ZIP: 19477
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/331,384

Query Match 0.9%; Score 37; DB 2; Length 4910;
Best Local Similarity 100.0%; Pred. No. 2.1e-07;
Matches 37; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 AAGCTTCATGCTCGAGGTCGACTCTAGAGGATCCG 37
Db 399 AAGCTTCATGCTCGAGGTCGACTCTAGAGGATCCG 435

RESULT 8
US-09-246-320-2
; Sequence 2, Application US/09246320
; Patent No. 6251677
; GENERAL INFORMATION:
; APPLICANT: Trustees of University of Pennsylvania
; APPLICANT: Wilson, James M.
; APPLICANT: Kelley, William M.
; APPLICANT: Fisher, Krishna J.
; TITLE OF INVENTION: Hybrid Adenovirus-AAV Vector and
; TITLE OF INVENTION: Methods of Use Thereof
; NUMBER OF SEQUENCES: 2
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Howson and Howson
; STREET: Spring House Corporate Cntr, PO Box 457
; CITY: Spring House
; STATE: Pennsylvania
; COUNTRY: USA
; ZIP: 19477
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent in Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/246,320
; FILING DATE:
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/836,087
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: Bak, Mary E.
; REGISTRATION NUMBER: 31,215
; REFERENCE/DOCKET NUMBER: GNPVN.007PCT
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 215-540-9200
; TELEFAX: 215-540-5818
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 4910 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: unknown
; MOLECULE TYPE: CDNA
US-08-331-384-2

Query Match 0.9%; Score 37; DB 2; Length 4910;
Best Local Similarity 100.0%; Pred. No. 2.1e-07;
Matches 37; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 AAGCTTCATGCTCGAGGTCGACTCTAGAGGATCCG 37
Db 399 AAGCTTCATGCTCGAGGTCGACTCTAGAGGATCCG 435

RESULT 7
US-08-836-087-2
; Sequence 2, Application US/08836087
; Patent No. 5871982
; GENERAL INFORMATION:
; APPLICANT: Trustees of University of Pennsylvania
; APPLICANT: Wilson, James M.
; APPLICANT: Kelley, William M.
; APPLICANT: Fisher, Krishna J.
; TITLE OF INVENTION: Hybrid Adenovirus-AAV Vector and
; TITLE OF INVENTION: Methods of Use Thereof
; NUMBER OF SEQUENCES: 2
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Howson and Howson
; STREET: Spring House Corporate Cntr, PO Box 457
; CITY: Spring House
; STATE: Pennsylvania
; COUNTRY: USA
; ZIP: 19477
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent in Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/836,087
; FILING DATE:
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/331,384
; FILING DATE: 28-OCT-1994
; ATTORNEY/AGENT INFORMATION:
; NAME: Bak, Mary E.
; REGISTRATION NUMBER: 31,215
; REFERENCE/DOCKET NUMBER: GNPVN.007PCT
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 215-540-9200
; TELEFAX: 215-540-5818
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 4910 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: unknown
; MOLECULE TYPE: CDNA
US-08-836-087-2

Query Match 0.9%; Score 37; DB 2; Length 4910;
Best Local Similarity 100.0%; Pred. No. 2.1e-07;
Matches 37; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 AAGCTTCATGCTCGAGGTCGACTCTAGAGGATCCG 37
Db 399 AAGCTTCATGCTCGAGGTCGACTCTAGAGGATCCG 435

RESULT 8
US-09-246-320-2
; Sequence 2, Application US/09246320
; Patent No. 6251677
; GENERAL INFORMATION:
; APPLICANT: Trustees of University of Pennsylvania
; APPLICANT: Wilson, James M.
; APPLICANT: Kelley, William M.
; APPLICANT: Fisher, Krishna J.
; TITLE OF INVENTION: Hybrid Adenovirus-AAV Vector and
; TITLE OF INVENTION: Methods of Use Thereof
; NUMBER OF SEQUENCES: 2
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Howson and Howson
; STREET: Spring House Corporate Cntr, PO Box 457
; CITY: Spring House
; STATE: Pennsylvania
; COUNTRY: USA
; ZIP: 19477
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent in Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/246,320
; FILING DATE:
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/836,087
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: Bak, Mary E.
; REGISTRATION NUMBER: 31,215
; REFERENCE/DOCKET NUMBER: GNPVN.007PCT
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 215-540-9200
; TELEFAX: 215-540-5818
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 4910 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: unknown
; MOLECULE TYPE: CDNA
US-09-246-320-2

Query Match 0.9%; Score 37; DB 4; Length 4910;
Best Local Similarity 100.0%; Pred. No. 2.1e-07;
Matches 37; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 AAGCTTCATGCTCGAGGTCGACTCTAGAGGATCCG 37
Db 399 AAGCTTCATGCTCGAGGTCGACTCTAGAGGATCCG 435

RESULT 9
US-09-546-738-2
; Sequence 2, Application US/09546738
; Patent No. 6387368
; GENERAL INFORMATION:
; APPLICANT: Trustees of University of Pennsylvania
; APPLICANT: Wilson, James M.

;; Kelley, William M.
;; Fisher, Krishna J.
;; TITLE OF INVENTION: Hybrid Adenovirus-AAV Vector and
;; Methods of Use Thereof
;;
;; NUMBER OF SEQUENCES: 2
;; CORRESPONDENCE ADDRESS:
;; ADDRESSEE: Howson and Howson
;; STREET: Spring House Corporate Cntr, PO Box 457
;; CITY: Spring House
;; STATE: Pennsylvania
;; COUNTRY: USA
;; ZIP: 19477
;;
;; COMPUTER READABLE FORM:
;; MEDIUM TYPE: Floppy disk
;; COMPUTER: IBM PC compatible
;; OPERATING SYSTEM: PC-DOS/MS-DOS
;; SOFTWARE: Patentin Release #1.0, Version #1.25
;;
;; CURRENT APPLICATION DATA:
;; APPLICATION NUMBER: US/09/546,738
;; FILING DATE: 11-Apr-2000
;; CLASSIFICATION: <Unknown>
;;
;; PRIOR APPLICATION DATA:
;; APPLICATION NUMBER: 09/246,320
;; FILING DATE: <Unknown>
;;
;; ATTORNEY/AGENT INFORMATION:
;; NAME: Bak, Mary E.
;; REGISTRATION NUMBER: 31,215
;; REFERENCE/DOCKET NUMBER: GNVPN.007PCT
;; TELECOMMUNICATION INFORMATION:
;; TELEPHONE: 215-540-9200
;; TELEFAX: 215-540-5818
;;
;; INFORMATION FOR SEQ ID NO: 2:
;; SEQUENCE CHARACTERISTICS:
;; LENGTH: 4910 base pairs
;; TYPE: nucleic acid
;; STRANDEDNESS: double
;; TOPOLOGY: unknown
;;
;; MOLECULE TYPE: CDNA
;; SEQUENCE DESCRIPTION: SEQ ID NO: 2:
US-09-546-738-2

Query Match 0.9%; Score 37; DB 4; Length 4910;
Best Local Similarity 100.0%; Pred. No. 2.le-07;
Matches 37; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 AAGCTTCATGCCCTGCAGGTCGACTCTAGAGGATCCG 37
|||||
Db 399 AAGCTTCATGCCCTGCAGGTCGACTCTAGAGGATCCG 435

RESULT 10
US-08-659-206A-4/c
; Sequence 4, Application US/08659206A
; Patent No. 5922685
; GENERAL INFORMATION:
; APPLICANT: Rakhmievich, Alexander
; TITLE OF INVENTION: IL-12 Gene Therapy of Tumors
; NUMBER OF SEQUENCES: 6
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Quarles & Brady
; STREET: 1 South Pinckney Street
; CITY: Madison
; STATE: WI
; COUNTRY: US
; ZIP: 53703
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/659,206A
; FILING DATE:

;; CLASSIFICATION: 435
;; ATTORNEY/AGENT INFORMATION:
;; NAME: Seay, Nicholas J.
;; REGISTRATION NUMBER: 27386
;; REFERENCE/DOCKET NUMBER: 110229.91144
;; TELECOMMUNICATION INFORMATION:
;; TELEPHONE: 608-251-5000
;; TELEFAX: 608-251-9166
;; INFORMATION FOR SEQ ID NO: 4:
;; SEQUENCE CHARACTERISTICS:
;; LENGTH: 6295 base pairs
;; TYPE: nucleic acid
;; STRANDEDNESS: double
;; TOPOLOGY: circular
;; MOLECULE TYPE: other nucleic acid
;; DESCRIPTION: /desc = "plasmid pWRG3196"
;; FEATURE:
;; NAME/KEY: CDS
;; LOCATION: join(955..1260, 1334..1675)
;; OTHER INFORMATION: /product= "p35 gene product"
;; FEATURE:
;; NAME/KEY: CDS
;; LOCATION: 2377..3384
;; OTHER INFORMATION: /product= "p40 gene product"
;;
;; US-08-659-206A-4

Query Match 0.9%; Score 37; DB 2; Length 6295;
Best Local Similarity 100.0%; Pred. No. 2.le-07;
Matches 37; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 AAGCTTCATGCCCTGCAGGTCGACTCTAGAGGATCCG 37
|||||
Db 3666 AAGCTTCATGCCCTGCAGGTCGACTCTAGAGGATCCG 3630

RESULT 11
US-08-836-022A-1
; Sequence 1, Application US/08836022A
; Patent No. 6001557
; GENERAL INFORMATION:
; APPLICANT: Trustees of the University of Pennsylvania
; APPLICANT: Wilson, James M.
; APPLICANT: Fisher, Krishna J.
; APPLICANT: Chen, Shu-Jen
; APPLICANT: Weitzman, Matthew
; TITLE OF INVENTION: Improved Adenovirus Virus and
; NUMBER OF SEQUENCES: 10
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Howson and Howson
; STREET: Spring House Corporate Cntr, P O Box 457
; CITY: Spring House
; STATE: Pennsylvania
; COUNTRY: USA
; ZIP: 19477
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/836,022A
; FILING DATE:
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/331,381
; FILING DATE: 28-OCT-1994
; ATTORNEY/AGENT INFORMATION:
; NAME: Bak, Mary E.
; REGISTRATION NUMBER: 31,215
; REFERENCE/DOCKET NUMBER: GNVPN.008PCT
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 215-540-9200
; TELEFAX: 215-540-5818

INFORMATION FOR SEQ ID NO: 1:
SEQUENCE CHARACTERISTICS:
LENGTH: 7897 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: unknown
MOLECULE TYPE: cDNA
US-08-836-022A-1

Query Match 0.9%; Score 37; DB 3; Length 7897;
Best Local Similarity 100.0%; Pred. No. 2e-07;
Matches 37; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 AAGCTTGCATCGCTGCAGTCTAGAGGATCCG 37
Db 609 AAGCTTGCATCGCTGCAGTCTAGAGGATCCG 645

RESULT 12
US-09-427-048A-1
Sequence 1, Application US/09427048A
Patent No. 6203975

GENERAL INFORMATION:
APPLICANT: Trustees of the University of Pennsylvania
Wilson, James M.
Fisher, Krishna J.
Chen, Shu-Jen

WEITZMAN, Matthew
TITLE OF INVENTION: Improved Adenovirus Virus and
Methods of Use Thereof

NUMBER OF SEQUENCES: 10
CORRESPONDENCE ADDRESS:
ADDRESSEE: Howson and Howson
STREET: Spring House Corporate Cntr, P O Box 457
CITY: Spring House
STATE: Pennsylvania
COUNTRY: USA
ZIP: 19477

COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent In Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/427,048A
FILING DATE: 21-Oct-1999
CLASSIFICATION: <unknown>

PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/836,022
FILING DATE: <unknown>
ATTORNEY/AGENT INFORMATION:
NAME: Bak, Mary E.
REGISTRATION NUMBER: 31,215
REFERENCE/DOCKET NUMBER: GNPVN.008PCT
TELECOMMUNICATION INFORMATION:
TELEPHONE: 215-540-9200
TELEFAX: 215-540-5818

INFORMATION FOR SEQ ID NO: 1:
SEQUENCE CHARACTERISTICS:
LENGTH: 7897 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: unknown
MOLECULE TYPE: cDNA
SEQUENCE DESCRIPTION: SEQ ID NO: 1:
US-09-427-048A-1

Query Match 0.9%; Score 37; DB 4; Length 7897;
Best Local Similarity 100.0%; Pred. No. 2e-07;
Matches 37; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 AAGCTTGCATCGCTGCAGTCTAGAGGATCCG 37
Db 1 AAGCTTGCATCGCTGCAGTCTAGAGGATCCG 37

Db 609 AAGCTTGCATCGCTGCAGTCTAGAGGATCCG 645

RESULT 13

US-08-487-283A-4/c
Sequence 4, Application US/08487283A
Patent No. 6355245

GENERAL INFORMATION:

APPLICANT: Evans, Mark J.
APPLICANT: Matis, Louis A.
APPLICANT: Mueller, Eileen Elliott
APPLICANT: Nye, Steven H.
APPLICANT: Rollins, Scott
APPLICANT: Rother, Russell P.
APPLICANT: Springhorn, Jeremy P.
APPLICANT: Squinto, Stephen P.
APPLICANT: Thomas, Thomas C.
APPLICANT: Wilkins, James A.

TITLE OF INVENTION: METHODS AND COMPOSITIONS FOR THE TREATMENT
OF INFLAMMATORY DISEASES
NUMBER OF SEQUENCES: 26
CORRESPONDENCE ADDRESS:
ADDRESSEE: Seth A. Fidel
STREET: 25 Science Park (Alexion)
CITY: New Haven
STATE: Connecticut
COUNTRY: USA
ZIP: 06511

COMPUTER READABLE FORM:
MEDIUM TYPE: 3.5 inch, 1.4Mb storage
COMPUTER: Macintosh Cetrus 610
OPERATING SYSTEM: System 7
SOFTWARE: WordPerfect 3.0
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/487,283A
FILING DATE: June 7, 1995
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/236,208
FILING DATE: 02-MAY-1994
ATTORNEY/AGENT INFORMATION:
NAME: Seth A. Fidel

REGISTRATION NUMBER: 38,449
REFERENCE/DOCKET NUMBER: ALX-152.1 CIP
TELECOMMUNICATION INFORMATION:
TELEPHONE: (203)776-1790
TELEFAX: (203)772-3655
INFORMATION FOR SEQ ID NO: 4:
SEQUENCE CHARACTERISTICS:
LENGTH: 8540 base pairs
TYPE: Nucleic Acid
STRANDEDNESS: Double
TOPOLOGY: Circular
MOLECULE TYPE: Other nucleic acid
DESCRIPTION: Apex-3P Eukaryotic
DESCRIPTION: Expression Vector
US-08-487-283A-4

Query Match 0.9%; Score 37; DB 4; Length 8540;
Best Local Similarity 100.0%; Pred. No. 2e-07;
Matches 37; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 AAGCTTGCATCGCTGCAGTCTAGAGGATCCG 37
Db 1371 AAGCTTGCATCGCTGCAGTCTAGAGGATCCG 1335

RESULT 14

PCT-US96-03611A-12/c

Sequence 12, Application PC/TUS9605611A
GENERAL INFORMATION:
APPLICANT: Mueller, John P.
APPLICANT: Lenardo, Michael J.
APPLICANT: McFarland, Henry F.

```

; APPLICANT: Matis, Louis A.
; APPLICANT: Mueller, Eileen Elliott
; APPLICANT: Nye, Steven H.
; APPLICANT: Pelfrey, Clara M.
; APPLICANT: Squinto, Stephen P.
; APPLICANT: Wilkins, James A.
; TITLE OF INVENTION: Modified Myelin Protein Molecules
; NUMBER OF SEQUENCES: 29
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Maurice M. Klee
; STREET: 1951 Burr Street
; CITY: Fairfield
; STATE: Connecticut
; COUNTRY: USA
; ZIP: 06430
; COMPUTER READABLE FORM:
; MEDIUM TYPE: 3.5 inch, 0.8 Mb storage
; COMPUTER: Macintosh Centris 610
; OPERATING SYSTEM: System 7
; SOFTWARE: Microsoft Word 6.0.1
; CURRENT APPLICATION NUMBER: PCT/US96/05611A
; APPLICATION NUMBER: PCT/US96/05611A
; FILING DATE: 02-MAY-1995
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/431,644
; FILING DATE: May 2, 1995
; APPLICATION NUMBER: 08/431,648
; FILING DATE: May 2, 1995
; APPLICATION NUMBER: 08/482,114
; FILING DATE: June 7, 1995
; ATTORNEY/AGENT INFORMATION:
; NAME: Klee, Maurice M.
; REGISTRATION NUMBER: 30,399
; REFERENCE/DOCKET NUMBER: ALX-129
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (203) 255 1400
; TELEFAX: (203) 254 1101
; INFORMATION FOR SEQ ID NO: 12:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 8540 base pairs
; TYPE: Nucleic Acid
; STRANDEDNESS: Double
; TOPOLOGY: Circular
; MOLECULE TYPE: Other nucleic acid
; DESCRIPTION: Apex-3P Eukaryotic
; DESCRIPTION: Expression Vector
; PCT-US96-05611A-12

Query Match 0.9%; Score 37; DB 5; Length 8540;
Best Local Similarity 100.0%; Pred. No. 2e-07;
Matches 37; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 AAGCTTGCATGCTGCAGGTGCGACTCTAGAGGATCCG 37
Db 1371 AAGCTTGCATGCTGCAGGTGCGACTCTAGAGGATCCG 1335

RESULT 15
US-08-252-493C-8/C
; Sequence 8, Application US/08252493C
; Patent No. 5891645
; GENERAL INFORMATION:
; APPLICANT: Rollins, Scott
; APPLICANT: Rother, Russell P.
; APPLICANT: Evans, Mark J.
; APPLICANT: Matis, Louis A.
; TITLE OF INVENTION: PORCINE E-SELECTIN
; NUMBER OF SEQUENCES: 9
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Seth A. Fidel
; STREET: 25 Science Park, Box 15
; CITY: New Haven

```

```

; STATE: Connecticut
; COUNTRY: USA
; ZIP: 06511
; COMPUTER READABLE FORM:
; MEDIUM TYPE: 3.5 inch, 750 Kb storage
; COMPUTER: PC compatible
; OPERATING SYSTEM: DOS 6.2
; SOFTWARE: WordPerfect 6.0
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/252,493C
; FILING DATE: June 1, 1994
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER:
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: Fidel, Seth A.
; REGISTRATION NUMBER: 38,449
; REFERENCE/DOCKET NUMBER: ALX-138
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (203) 776-1790
; TELEFAX: (203) 772-3655
; INFORMATION FOR SEQ ID NO: 8:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 8932 base pairs
; TYPE: Nucleic Acid
; STRANDEDNESS: Double
; TOPOLOGY: Circular
; MOLECULE TYPE: Other nucleic acid
; DESCRIPTION: Apex-3 Eukaryotic
; DESCRIPTION: Expression Vector
; US-08-252-493C-8

Query Match 0.9%; Score 37; DB 2; Length 8932;
Best Local Similarity 100.0%; Pred. No. 2e-07;
Matches 37; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 AAGCTTGCATGCTGCAGGTGCGACTCTAGAGGATCCG 37
Db 1371 AAGCTTGCATGCTGCAGGTGCGACTCTAGAGGATCCG 1335

Search completed: May 19, 2003, 01:51:55
Job time : 235 secs

```

THIS PAGE BLANK (USPTO)

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: May 19, 2003, 01:48:05 ; Search time 457 Seconds
(without alignments)
11615.089 Million cell updates/sec

Title: US-09-926-163B-1
Perfect score: 4115
Sequence: 1 aagcttgatgctgcaggt.....tccggcgccgctgaagctt 4115

Scoring table: OIIGO_NUC
Gapop 60.0 , Gapext 60.0

Searched: 810007 seqs, 644969091 residues

Word size : 0

Total number of hits satisfying chosen parameters: 1620014

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Database : Published Applications,NA:*

- 1: /cgn2_6/ptodata/2/pubpna/US07_PUBCOMB.seq:*
- 2: /cgn2_6/ptodata/2/pubpna/PCT_NEW_PUB.seq:*
- 3: /cgn2_6/ptodata/2/pubpna/US06_NEW_PUB.seq:*
- 4: /cgn2_6/ptodata/2/pubpna/US06_PUBCOMB.seq:*
- 5: /cgn2_6/ptodata/2/pubpna/US07_NEW_PUB.seq:*
- 6: /cgn2_6/ptodata/2/pubpna/PCTUS_PUBCOMB.seq:*
- 7: /cgn2_6/ptodata/2/pubpna/US08_NEW_PUB.seq:*
- 8: /cgn2_6/ptodata/2/pubpna/US08_PUBCOMB.seq:*
- 9: /cgn2_6/ptodata/2/pubpna/US09_NEW_PUB.seq:*
- 10: /cgn2_6/ptodata/2/pubpna/US09_PUBCOMB.seq:*
- 11: /cgn2_6/ptodata/2/pubpna/US10_NEW_PUB.seq:*
- 12: /cgn2_6/ptodata/2/pubpna/US10_PUBCOMB.seq:*
- 13: /cgn2_6/ptodata/2/pubpna/US60_NEW_PUB.seq:*
- 14: /cgn2_6/ptodata/2/pubpna/US60_PUBCOMB.seq:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Query Match	Score	Length	ID	Description
C 1	37	0.9	3796	9	US-10-127-391-32
C 2	36	0.9	46	9	US-09-940-925A-161
C 3	36	0.9	57	10	US-09-785-269-1
C 4	36	0.9	59	9	US-09-939-275A-3
C 5	36	0.9	59	12	US-10-021-237-5
C 6	36	0.9	64	9	US-09-940-925A-73
C 7	36	0.9	66	10	US-09-785-269-8
C 8	36	0.9	71	10	US-09-380-932-3
C 9	36	0.9	75	9	US-09-972-985-6
C 10	36	0.9	75	10	US-09-955-649-6
C 11	36	0.9	75	10	US-09-973-013-6
C 12	36	0.9	89	9	US-10-135-807-11
C 13	36	0.9	102	9	US-10-015-219-106
C 14	36	0.9	102	10	US-09-777-564-106
C 15	36	0.9	103	10	US-09-179-536B-267
C 16	36	0.9	119	8	US-08-852-020-15
C 17	36	0.9	133	10	US-09-345-761-4
C 18	36	0.9	141	9	US-09-940-925A-162
C 19	36	0.9	149	12	US-10-012-211A-5

20	36	0.9	165	9	US-09-940-925A-9	Sequence 49, Appl
C 21	36	0.9	206	9	US-10-033-297-27	Sequence 27, Appl
C 22	36	0.9	206	9	US-10-081-806-32	Sequence 32, Appl
C 23	36	0.9	206	9	US-10-074-328-32	Sequence 32, Appl
C 24	36	0.9	206	9	US-09-940-244-27	Sequence 27, Appl
C 25	36	0.9	206	9	US-09-940-925A-32	Sequence 32, Appl
C 26	36	0.9	206	9	US-09-940-925A-50	Sequence 50, Appl
C 27	36	0.9	228	9	US-09-940-925A-165	Sequence 165, Appl
C 28	36	0.9	293	9	US-10-132-561-11	Sequence 11, Appl
C 29	36	0.9	293	9	US-10-132-561-12	Sequence 12, Appl
C 30	36	0.9	298	9	US-09-808-124A-4	Sequence 4, Appl
C 31	36	0.9	298	10	US-09-845-157-6	Sequence 6, Appl
C 32	36	0.9	355	9	US-09-910-009A-157	Sequence 157, Appl
C 33	36	0.9	384	10	US-09-757-992-6	Sequence 6, Appl
C 34	36	0.9	384	10	US-09-757-992-13	Sequence 13, Appl
C 35	36	0.9	440	10	US-09-878-574-4506	Sequence 4506, Ap
C 36	36	0.9	458	9	US-10-132-561-14	Sequence 14, Appl
C 37	36	0.9	458	9	US-10-132-561-13	Sequence 13, Appl
C 38	36	0.9	459	9	US-10-132-561-13	Sequence 13, Appl
C 39	36	0.9	498	10	US-09-894-882-53	Sequence 53, Appl
C 40	36	0.9	498	10	US-09-894-882-8	Sequence 8, Appl
C 41	36	0.9	507	10	US-09-772-134B-26	Sequence 26, Appl
C 42	36	0.9	560	10	US-09-772-134B-29	Sequence 29, Appl
C 43	36	0.9	566	10	US-10-052-154-1873	Sequence 1873, Ap
C 44	36	0.9	586	9	US-09-764-847-1873	Sequence 1873, Ap
C 45	36	0.9	586	10	US-09-764-847-1873	Sequence 1873, Ap

ALIGNMENTS

RESULT 1
US-10-127-391-32/C
; Sequence 32, Application US/10127391
; Patent No. US20020172690A1
; GENERAL INFORMATION:
; APPLICANT: Pfizer Products Inc.
; TITLE OF INVENTION: AN INFECTIOUS CDNA CLONE OF NORTH AMERICAN PORCINE
; TITLE OF INVENTION: REPRODUCTIVE AND RESPIRATORY SYNDROME (PRRS) VIRUS AND
; FILE OF INVENTION: US09-940-925A-73
; FILE REFERENCE: PC10278A
; CURRENT APPLICATION NUMBER: US/10/127,391
; CURRENT FILING DATE: 2002-04-22
; PRIOR APPLICATION NUMBER: US/09/470,661A
; PRIOR FILING DATE: 1999-12-22
; NUMBER OF SEQ ID NOS: 45
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 32
; LENGTH: 3796
; TYPE: DNA
; ORGANISM: Plasmid
; FEATURE:
; OTHER INFORMATION: Description of Plasmid: pCMV-MC1
US-10-127-391-32

Query Match 0.9%; Score 37; DB 9; Length 3796;
Best Local Similarity 100.0%; Pred. No. 1.6e-09;
Matches 37; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 AAGCTTGCATGCGCTGCAGGTCGACTCTAGAGGATCCG 37
|||||
Db 1167 AAGCTTGCATGCGCTGCAGGTCGACTCTAGAGGATCCG 1131

RESULT 2
US-09-940-925A-161
; Sequence 161, Application US/09940925A
; Publication No. US20030054338A1
; GENERAL INFORMATION:
; APPLICANT: BROW, MARY ANN D.
; LYAMICHEV, VICTOR I.
; OLIVE, DAVID M.
; TITLE OF INVENTION: RAPID DETECTION AND IDENTIFICATION OF

; PATHOGENS
;
; NUMBER OF SEQUENCES: 165
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: MEDLEN & CARROLL
; STREET: 220 MONTGOMERY STREET, SUITE 2200
; CITY: SAN FRANCISCO
; STATE: CALIFORNIA
; COUNTRY: UNITED STATES OF AMERICA
; ZIP: 94104
;
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.30
;
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/940,925A
; FILING DATE: 10-Jun-2002
; CLASSIFICATION: <Unknown>
; ATTORNEY/AGENT INFORMATION:
; NAME: CARROLL, PETER G.
; REGISTRATION NUMBER: 32,837
; REFERENCE/DOCKET NUMBER: FORS-01756
;
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (415) 705-8410
; TELEFAX: (415) 397-8338
;
; INFORMATION FOR SEQ ID NO: 161:
;
; SEQUENCE CHARACTERISTICS:
; LENGTH: 46 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: RNA (genomic)
; SEQUENCE DESCRIPTION: SEQ ID NO: 161:
US-09-940-925A-161

Query Match 0.9%; Score 36; DB 9; Length 46;
Best Local Similarity 77.8%; Pred. No. 7.3e-09;
Matches 28; Conservative 8; Mismatches 0; Indels 0; Gaps 0;

QY 1 AAGCTTGATCGCTGCAGGTGCGACTCTAGAGGATCC 36
|||||:|||||:|||||:|||||:|||||:|||||:|
DB 9 AAGCUUGAUGCCGCGAGGUGACUCUAGAGGAUCC 44

RESULT 3
US-09-785-269-1
; Sequence 1, Application US/09785269
; Patent No. US20020018736A1
; GENERAL INFORMATION:
; APPLICANT: IMAI, Kensaku
; KITAJIMA, Masato
;
; TITLE OF INVENTION: METHOD AND APPARATUS FOR AUTOMATICALLY
; REMOVING VECTOR UNIT IN DNA BASE SEQUENCE
;
; NUMBER OF SEQUENCES: 19
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Staas & Halsey
; STREET: 700 Eleventh Street, N.W., Suite 500
; CITY: Washington
; STATE: DC
; COUNTRY: US
; ZIP: 20001
;
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.30
;
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/785,269
; FILING DATE: 20-Feb-2001
; CLASSIFICATION: <Unknown>
;
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/684,674
; FILING DATE: 22-JUL-1996

; ATTORNEY/AGENT INFORMATION:
; NAME: Herbert, William F.
; REGISTRATION NUMBER: 31,024
; REFERENCE/DOCKET NUMBER: 862.1335/WEH
;
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 2024341500
; TELEFAX: 2024341501
;
; INFORMATION FOR SEQ ID NO: 1:
;
; SEQUENCE CHARACTERISTICS:
; LENGTH: 57
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
; SEQUENCE DESCRIPTION: SEQ ID NO: 1:
US-09-785-269-1

Query Match 0.9%; Score 36; DB 10; Length 57;
Best Local Similarity 100.0%; Pred. No. 7.2e-09;
Matches 36; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 AAGCTTGATCGCTGCAGGTGCGACTCTAGAGGATCC 36
|||||:|||||:|||||:|||||:|||||:|||||:|
DB 1 AAGCTTGATCGCTGCAGGTGCGACTCTAGAGGATCC 36

RESULT 4
US-09-939-275A-3/c
; Sequence 3, Application US/09939275A
; Patent No. US20020172955A1
; GENERAL INFORMATION:
; APPLICANT: Adams, Christopher
; APPLICANT: Boles, T. Christian
; APPLICANT: Weir, Lawrence
; APPLICANT: Dhanda, Rahul
; APPLICANT: Summers, Nevin
;
; TITLE OF INVENTION: Methods for Purifying DNA Using Immobilized Capture Probes
;
; FILE REFERENCE: EXT-062CN
; CURRENT APPLICATION NUMBER: US/09/939,275A
; CURRENT FILING DATE: 2001-08-24
; PRIOR APPLICATION NUMBER: US 09/259,467
; PRIOR FILING DATE: 1999-02-26
; PRIOR APPLICATION NUMBER: PCT/US98/09952
; PRIOR FILING DATE: 1998-05-15
; PRIOR APPLICATION NUMBER: US 08/971,845
; PRIOR FILING DATE: 1997-08-08
; PRIOR APPLICATION NUMBER: US 60/046,708
; PRIOR FILING DATE: 1997-05-16
; NUMBER OF SEQ ID NOS: 3
; SOFTWARE: Patent In version 3.0
; SEQ ID NO 3
; LENGTH: 59
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: template region
US-09-939-275A-3

Query Match 0.9%; Score 36; DB 9; Length 59;
Best Local Similarity 100.0%; Pred. No. 7.2e-09;
Matches 36; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 AAGCTTGATCGCTGCAGGTGCGACTCTAGAGGATCC 36
|||||:|||||:|||||:|||||:|||||:|||||:|
DB 37 AAGCTTGATCGCTGCAGGTGCGACTCTAGAGGATCC 2

RESULT 5
US-10-021-237-5/c
; Sequence 5, Application US/10021237
; Patent No. US20020119480A1
; GENERAL INFORMATION:
; APPLICANT: Weir, Lawrence

APPLICANT: Adams, Christopher
APPLICANT: Boles, T. Christian
APPLICANT: Dhanda, Rahul
APPLICANT: Kron, Stephen
TITLE OF INVENTION: Purification Devices Comprising Immobilized Capture Probes and Us
TITLE OF INVENTION: Therefor
FILE REFERENCE: EXT-070C1
CURRENT APPLICATION NUMBER: US/10/021,237
PRIOR FILING DATE: 2001-12-06
PRIOR APPLICATION NUMBER: US 09/513,381
PRIOR FILING DATE: 2000-02-25
PRIOR APPLICATION NUMBER: US 60/121,836
PRIOR FILING DATE: 1999-02-26
NUMBER OF SEQ ID NOS: 9
SOFTWARE: PatentIn version 3.0
SEQ ID NO 5
LENGTH: 59
TYPE: DNA
ORGANISM: artificial sequence
FEATURE:
OTHER INFORMATION: polylinker.
US-10-021-237-5

Query Match 0.9% Score 36; DB 12; Length 59;
Best Local Similarity 100.0%; Pred. No. 7.2e-09;
Matches 36; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 AAGCTTGCATGCGCTGCAGGTCGACTCTAGAGGATCC 36
Db 37 AAGCTTGCATGCGCTGCAGGTCGACTCTAGAGGATCC 2

RESULT 6
US-09-940-925A-73
Sequence 73, Application US/09940925A
Publication No. US20030054338A1
GENERAL INFORMATION:
APPLICANT: BROW, MARY ANN D.
LYAMICHEV, VICTOR I.
OLIVE, DAVID M.
TITLE OF INVENTION: RAPID DETECTION AND IDENTIFICATION OF
PATHOGENS
NUMBER OF SEQUENCES: 165
CORRESPONDENCE ADDRESS:
ADDRESSEE: MEDLEN & CARROLL
STREET: 220 MONTGOMERY STREET, SUITE 2200
CITY: SAN FRANCISCO
STATE: CALIFORNIA
COUNTRY: UNITED STATES OF AMERICA
ZIP: 94104
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/940,925A
FILING DATE: 10-Jun-2002
CLASSIFICATION: <Unknown>
ATTORNEY/AGENT INFORMATION:
NAME: CARROLL, PETER G.
REGISTRATION NUMBER: 32,837
REFERENCE/DOCKET NUMBER: FORS-01756
TELECOMMUNICATION INFORMATION:
TELEPHONE: (415) 705-8410
TELEFAX: (415) 397-8338
INFORMATION FOR SEQ ID NO: 73:
SEQUENCE CHARACTERISTICS:
LENGTH: 64 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: RNA (genomic)

APPLICANT: Adams, Christopher
APPLICANT: Boles, T. Christian
APPLICANT: Dhanda, Rahul
APPLICANT: Kron, Stephen
TITLE OF INVENTION: Purification Devices Comprising Immobilized Capture Probes and Us
TITLE OF INVENTION: Therefor
FILE REFERENCE: EXT-070C1
CURRENT APPLICATION NUMBER: US/10/021,237
PRIOR FILING DATE: 2001-12-06
PRIOR APPLICATION NUMBER: US 09/513,381
PRIOR FILING DATE: 2000-02-25
PRIOR APPLICATION NUMBER: US 60/121,836
PRIOR FILING DATE: 1999-02-26
NUMBER OF SEQ ID NOS: 9
SOFTWARE: PatentIn version 3.0
SEQ ID NO 5
LENGTH: 59
TYPE: DNA
ORGANISM: artificial sequence
FEATURE:
OTHER INFORMATION: polylinker.
US-10-021-237-5

Query Match 0.9% Score 36; DB 12; Length 59;
Best Local Similarity 100.0%; Pred. No. 7.2e-09;
Matches 36; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 AAGCTTGCATGCGCTGCAGGTCGACTCTAGAGGATCC 36
Db 37 AAGCTTGCATGCGCTGCAGGTCGACTCTAGAGGATCC 2

RESULT 6
US-09-940-925A-73
Sequence 73, Application US/09940925A
Publication No. US20030054338A1
GENERAL INFORMATION:
APPLICANT: BROW, MARY ANN D.
LYAMICHEV, VICTOR I.
OLIVE, DAVID M.
TITLE OF INVENTION: RAPID DETECTION AND IDENTIFICATION OF
PATHOGENS
NUMBER OF SEQUENCES: 165
CORRESPONDENCE ADDRESS:
ADDRESSEE: MEDLEN & CARROLL
STREET: 220 MONTGOMERY STREET, SUITE 2200
CITY: SAN FRANCISCO
STATE: CALIFORNIA
COUNTRY: UNITED STATES OF AMERICA
ZIP: 94104
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/940,925A
FILING DATE: 10-Jun-2002
CLASSIFICATION: <Unknown>
ATTORNEY/AGENT INFORMATION:
NAME: CARROLL, PETER G.
REGISTRATION NUMBER: 32,837
REFERENCE/DOCKET NUMBER: FORS-01756
TELECOMMUNICATION INFORMATION:
TELEPHONE: (415) 705-8410
TELEFAX: (415) 397-8338
INFORMATION FOR SEQ ID NO: 73:
SEQUENCE CHARACTERISTICS:
LENGTH: 64 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: RNA (genomic)

SEQUENCE DESCRIPTION: SEQ ID NO: 73:
US-09-940-925A-73

Query Match 0.9% Score 36; DB 9; Length 64;
Best Local Similarity 77.8%; Pred. No. 7.2e-09;
Matches 28; Conservative 8; Mismatches 0; Indels 0; Gaps 0;

Qy 1 AAGCTTGCATGCGCTGCAGGTCGACTCTAGAGGATCC 36
Db 9 AAGCUUGCAUGCCUGCAGGUGGACUCUAGAGGAUCC 44

RESULT 7
US-09-785-269-8
Sequence 8, Application US/09785269
Patent No. US20020018736A1
GENERAL INFORMATION:
APPLICANT: IMAI, Kensaku
KITAJIMA, Masato
TITLE OF INVENTION: METHOD AND APPARATUS FOR AUTOMATICALLY
REMOVING VECTOR UNIT IN DNA BASE SEQUENCE
NUMBER OF SEQUENCES: 19
CORRESPONDENCE ADDRESS:
ADDRESSEE: Staas & Halsey
STREET: 700 Eleventh Street, N.W., Suite 500
CITY: Washington
STATE: DC
COUNTRY: US
ZIP: 20001
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/785,269
FILING DATE: 20-Feb-2001
CLASSIFICATION: <Unknown>
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/684,674
FILING DATE: 22-JUL-1996
ATTORNEY/AGENT INFORMATION:
NAME: Herbert, William F.
REGISTRATION NUMBER: 31,024
REFERENCE/DOCKET NUMBER: 862.1335/WFH
TELECOMMUNICATION INFORMATION:
TELEPHONE: 2024341500
TELEFAX: 2024341501
INFORMATION FOR SEQ ID NO: 8:
SEQUENCE CHARACTERISTICS:
LENGTH: 66
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
SEQUENCE DESCRIPTION: SEQ ID NO: 8:
US-09-785-269-8

Query Match 0.9% Score 36; DB 10; Length 66;
Best Local Similarity 100.0%; Pred. No. 7.2e-09;
Matches 36; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 AAGCTTGCATGCGCTGCAGGTCGACTCTAGAGGATCC 36
Db 6 AAGCTTGCATGCGCTGCAGGTCGACTCTAGAGGATCC 41

RESULT 8
US-09-380-932-3
Sequence 3, Application US/09380932
Patent No. US20020058250A1
GENERAL INFORMATION:
APPLICANT: FIRTH, Greg

; TITLE OF INVENTION: EXTRACTION AND UTILISATION OF VNTR ALLELES
; FILE REFERENCE: 28911/35930
; CURRENT APPLICATION NUMBER: US/09/380,932
; CURRENT FILING DATE: 1999-03-21
; PRIOR APPLICATION NUMBER: PCT/GB98/00840
; PRIOR FILING DATE: 1998-03-20
; PRIOR APPLICATION NUMBER: EP 97301917.7
; PRIOR FILING DATE: 1997-03-21
; NUMBER OF SEQ ID NOS: 10
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 3
; LENGTH: 71
; TYPE: DNA
; ORGANISM: genomic DNA
US-09-380-932-3

Query Match 0.9%; Score 36; DB 10; Length 71;
Best Local Similarity 100.0%; Pred. No. 7.2e-09;
Matches 36; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
Qy 1 AAGCTTGATCGCTCGAGTGCAGTCTAGAGGATCC 36
|||||
Db 4 AAGCTTGATCGCTCGAGTGCAGTCTAGAGGATCC 39

RESULT 9

US-09-972-985-6
; Sequence 6, Application US/09972985
; Publication No. US20030066395A1
; GENERAL INFORMATION:
; APPLICANT: Ladant, Daniel
; APPLICANT: Karimova, Agnes
; APPLICANT: Ullmann, Agnes
; TITLE OF INVENTION: BACTERIAL MULTI-HYBRID SYSTEM AND APPLICATIONS
; TITLE OF INVENTION: THEREOF
; FILE REFERENCE: 03495-0178
; CURRENT APPLICATION NUMBER: US/09/972,985
; CURRENT FILING DATE: 2001-10-10
; PRIOR APPLICATION NUMBER: 09/203,681
; PRIOR FILING DATE: 1998-12-01
; NUMBER OF SEQ ID NOS: 11
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 6
; LENGTH: 75
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: Synthetic DNA
; NAME/KEY: CDS
; LOCATION: (1)..(75)
US-09-972-985-6

Query Match 0.9%; Score 36; DB 9; Length 75;
Best Local Similarity 100.0%; Pred. No. 7.1e-09;
Matches 36; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
Qy 1 AAGCTTGATCGCTCGAGTGCAGTCTAGAGGATCC 36
|||||
Db 18 AAGCTTGATCGCTCGAGTGCAGTCTAGAGGATCC 53

RESULT 10

US-09-955-649-6
; Sequence 6, Application US/09955649
; Patent No. US20020033812A1
; GENERAL INFORMATION:
; APPLICANT: Ladant, Daniel
; APPLICANT: Karimova, Gouzel
; APPLICANT: Ullmann, Agnes
; TITLE OF INVENTION: BACTERIAL MULTI-HYBRID SYSTEM AND APPLICATIONS THEREOF
; FILE REFERENCE: 03495-0190
; CURRENT APPLICATION NUMBER: US/09/955,649
; CURRENT FILING DATE: 2000-08-01

; PRIOR APPLICATION NUMBER: PCT/IB98/02085
; PRIOR FILING DATE: 1998-12-04
; NUMBER OF SEQ ID NOS: 11
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 6
; LENGTH: 75
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: Synthetic DNA
; NAME/KEY: CDS
; LOCATION: (1)..(75)
US-09-955-649-6

Query Match 0.9%; Score 36; DB 10; Length 75;
Best Local Similarity 100.0%; Pred. No. 7.1e-09;
Matches 36; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
Qy 1 AAGCTTGATCGCTCGAGTGCAGTCTAGAGGATCC 36
|||||
Db 18 AAGCTTGATCGCTCGAGTGCAGTCTAGAGGATCC 53

RESULT 11

US-09-973-013-6
; Sequence 6, Application US/09973013
; Patent No. US20020106783A1
; GENERAL INFORMATION:
; APPLICANT: Ladant, Daniel
; APPLICANT: Karimova, Gouzel
; APPLICANT: Ullmann, Agnes
; TITLE OF INVENTION: BACTERIAL MULTI-HYBRID SYSTEM AND APPLICATIONS THEREOF
; FILE REFERENCE: 03495-0178-01
; CURRENT APPLICATION NUMBER: US/09/973,013
; CURRENT FILING DATE: 2002-01-09
; PRIOR APPLICATION NUMBER: US 09/203,681
; PRIOR FILING DATE: 1998-12-01
; NUMBER OF SEQ ID NOS: 11
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 6
; LENGTH: 75
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: Synthetic DNA
; NAME/KEY: CDS
; LOCATION: (1)..(75)
US-09-973-013-6

Query Match 0.9%; Score 36; DB 10; Length 75;
Best Local Similarity 100.0%; Pred. No. 7.1e-09;
Matches 36; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
Qy 1 AAGCTTGATCGCTCGAGTGCAGTCTAGAGGATCC 36
|||||
Db 18 AAGCTTGATCGCTCGAGTGCAGTCTAGAGGATCC 53

RESULT 12

US-10-135-807-11/c
; Sequence 11, Application US/10135807
; Publication No. US20030049655A1
; GENERAL INFORMATION:
; APPLICANT: The Secretary of State for Defence in Her Britannic Majesty's
; APPLICANT: Government of the United Kingdom of Great Britain and
; APPLICANT: No. US20030049655A1
; APPLICANT: Vincent, Suzanne P
; APPLICANT: Clark, Duncan Roy
; TITLE OF INVENTION: Amplification process
; FILE REFERENCE: CG/P/133/WOD
; CURRENT APPLICATION NUMBER: US/10/135,807
; CURRENT FILING DATE: 2002-04-30
; PRIOR APPLICATION NUMBER: GB 0110501.4

```

; PRIOR FILING DATE: 2001-04-30
; NUMBER OF SEQ ID NOS: 26
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 11
; LENGTH: 89
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: Modified
; OTHER INFORMATION: polylinker sequence of pTQ18NHK
US-10-135-807-11

Query Match          0.9%; Score 36; DB 9; Length 89;
Best Local Similarity 100.0%; Pred. No. 7.1e-09;
Matches 36; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 AAGCTTGATCGCTGCAGGTGCTAGAGGATCC 36
    |||||||
Db 89 AAGCTTGATCGCTGCAGGTGCTAGAGGATCC 54
    |||||||

RESULT 13
US-10-015-219-106/c
; Sequence 106, Application US/10015219
; Publication No. US2003008299A1
; GENERAL INFORMATION:
; APPLICANT: Algate, Paul A.
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE THERAPY
; FILE REFERENCE: 210121.493C1
; CURRENT APPLICATION NUMBER: US/10/015,219
; CURRENT FILING DATE: 2002-03-02
; NUMBER OF SEQ ID NOS: 1739
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 106
; LENGTH: 102
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: 2
; OTHER INFORMATION: n = A,T,C or G
US-10-015-219-106

Query Match          0.9%; Score 36; DB 9; Length 102;
Best Local Similarity 100.0%; Pred. No. 7e-09;
Matches 36; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 AAGCTTGATCGCTGCAGGTGCTAGAGGATCC 36
    |||||||
Db 77 AAGCTTGATCGCTGCAGGTGCTAGAGGATCC 42

RESULT 14
US-09-777-564-106/c
; Sequence 106, Application US/09777564
; Patent No. US20020022591A1
; GENERAL INFORMATION:
; APPLICANT: Mannion, Jane
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE THERAPY
; FILE REFERENCE: 210121.493
; CURRENT APPLICATION NUMBER: US/09/777,564
; CURRENT FILING DATE: 2001-02-05
; NUMBER OF SEQ ID NOS: 1730
; SOFTWARE: FastSeq for Window Version 4.0
; SEQ ID NO 106
; LENGTH: 102
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc_feature

```

```

; LOCATION: (1)...(102)
; OTHER INFORMATION: n = A,T,C or G
US-09-777-564-106

Query Match          0.9%; Score 36; DB 10; Length 102;
Best Local Similarity 100.0%; Pred. No. 7e-09;
Matches 36; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 AAGCTTGATCGCTGCAGGTGCTAGAGGATCC 36
    |||||||
Db 77 AAGCTTGATCGCTGCAGGTGCTAGAGGATCC 42

RESULT 15
US-09-179-536B-267
; Sequence 267, Application US/09179536B
; Patent No. US20020042112A1
; GENERAL INFORMATION:
; APPLICANT: Hubert K ster
; David M. Lough
; Guobing Xiang
; TITLE OF INVENTION: DNA DIAGNOSTICS BASED ON MASS SPECTROMETRY
; NUMBER OF SEQUENCES: 320
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Heller Ehrman White & McAuliffe
; STREET: 4250 Executive Square, 7th Floor
; CITY: La Jolla
; STATE: CA
; COUNTRY: USA
; ZIP: 92037
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: DOS
; SOFTWARE: ASCII
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/179,536B
; FILING DATE: 26-Oct-1998
; CLASSIFICATION: <Unknown>
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: PCT/US97/20444
; FILING DATE: 06-NOV-1997
; APPLICATION NUMBER: 08/947,801
; FILING DATE: 08-Oct-97
; APPLICATION NUMBER: 08/933,792
; FILING DATE: 19-Sep-97
; APPLICATION NUMBER: 08/787,639
; FILING DATE: 23-Jan-97
; APPLICATION NUMBER: 08/786,988
; FILING DATE: 23-Jan-97
; APPLICATION NUMBER: 08/746,055
; FILING DATE: 06-No. US20020042112A1-96
; APPLICATION NUMBER: 08/746,036
; FILING DATE: 06-No. US20020042112A1-96
; APPLICATION NUMBER: 08/744,590
; FILING DATE: 06-No. US20020042112A1-96
; APPLICATION NUMBER: 08/744,481
; FILING DATE: 06-No. US20020042112A1-96
; ATTORNEY/AGENT INFORMATION:
; NAME: Seidman, Stephanie L
; REGISTRATION NUMBER: 33,779
; REFERENCE/DOCKET NUMBER: 24736-2004B
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 858-450-8400
; TELEFAX: 858-587-5360
; TELEX: <Unknown>
; INFORMATION FOR SEQ ID NO: 267:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 103 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: unknown
; MOLECULE TYPE: cDNA

```

```

; HYPOTHETICAL: NO
; ANTI-SENSE: NO
; FRAGMENT TYPE: <Unknown>
; ORIGINAL SOURCE:
; SEQUENCE DESCRIPTION: SEQ ID NO: 267:
US-09-179-536B-267

```

	Query Match	0.9%	Score 36;	DB 10;	Length 103;
	Best Local Similarity	100.0%;	Pred. No. 7e-09;		
	Matches 36;	Conservative 0;	Mismatches 0;	Indels 0;	Gaps 0;
QY	1	AAGCTTGCATGCTCGAGCTCGACTCTAGAGGATCC	36		
Db	21	AAGCTTGCATGCTCGAGCTCGACTCTAGAGGATCC	56		

Search completed: May 19, 2003, 07:50:56
Job time : 461 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: May 19, 2003, 02:27:41 ; Search time 5382 Seconds
(without alignments)
12382.837 Million cell updates/sec

Title: US-09-926-163B-1

Perfect score: 4115

Sequence: 1 aagcttgcacgtcgaggt.....tccggcggcctgaagctt 4115

Scoring table: OLIGO_NUC

Gapop 60.0 , Gapext 60.0

Searched: 16154066 seqs, 8097743376 residues

Word size : 0

Total number of hits satisfying chosen parameters: 32308132

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Database : EST.*

- 1: em_estba.*
- 2: em_esthum.*
- 3: em_estin.*
- 4: em_estmu.*
- 5: em_estov.*
- 6: em_estpl.*
- 7: em_estro.*
- 8: em_htc.*
- 9: gb_est1.*
- 10: gb_est2.*
- 11: gb_hic.*
- 12: gb_est3.*
- 13: gb_est4.*
- 14: gb_est5.*
- 15: em_estfun.*
- 16: em_estom.*
- 17: gb_gss.*
- 18: em_gss_hum.*
- 19: em_gss_inv.*
- 20: em_gss_pin.*
- 21: em_gss_vrt.*
- 22: em_gss_fun.*
- 23: em_gss_mam.*
- 24: em_gss_mus.*
- 25: em_gss_othr.*
- 26: em_gss_pro.*
- 27: em_gss_rod.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	39	0.9	241	14	225777
c 2	38	0.9	534	17	B60048
c 3	38	0.9	617	17	B70879
c 4	38	0.9	672	17	B70567
c 5	38	0.9	739	17	B64148
c 6	38	0.9	763	17	AQ916288

c 7	37	0.9	63	17	B32178
c 8	37	0.9	123	17	B40882
c 9	37	0.9	161	17	B79682
c 10	37	0.9	205	14	BM851683
c 11	37	0.9	206	14	BM850221
c 12	37	0.9	207	17	CNS00271
c 13	37	0.9	210	17	B33347
c 14	37	0.9	218	9	AT000691
c 15	37	0.9	221	17	AQ006733
c 16	37	0.9	246	17	B31188
c 17	37	0.9	253	2	BM821080
c 18	37	0.9	255	2	BM740520
c 19	37	0.9	285	17	B31114
c 20	37	0.9	285	17	B99340
c 21	37	0.9	290	14	BM850232
c 22	37	0.9	294	17	B38698
c 23	37	0.9	296	17	AQ080154
c 24	37	0.9	301	17	AQ080152
c 25	37	0.9	304	17	B40812
c 26	37	0.9	306	17	AQ002527
c 27	37	0.9	308	17	B34609
c 28	37	0.9	309	17	B32855
c 29	37	0.9	309	17	B41250
c 30	37	0.9	313	14	BM851117
c 31	37	0.9	315	17	B32887
c 32	37	0.9	319	2	BM852732
c 33	37	0.9	321	2	BM850274
c 34	37	0.9	322	17	B31778
c 35	37	0.9	323	2	BM821456
c 36	37	0.9	324	14	BM821148
c 37	37	0.9	330	14	BM849662
c 38	37	0.9	331	2	BM852441
c 39	37	0.9	333	2	BM747927
c 40	37	0.9	334	2	BM829792
c 41	37	0.9	341	17	B30702
c 42	37	0.9	342	17	B32593
c 43	37	0.9	344	17	AQ059544
c 44	37	0.9	348	2	BM852504
c 45	37	0.9	348	17	B45032

ALIGNMENTS

RESULT 1
225777
LOCUS
DEFINITION
225777
ACCESSION
225777.1 GI:3702537
KEYWORDS
SOURCE
ORGANISM
Homo sapiens
human.
REFERENCE
AUTHORS
1 (bases 1 to 241)
S.V. Ermolaeva, O.D., Lavrentyeva, I., Monastyrskaya, G.S. and Sverdllov, E.D.
TITLE
JOURNAL
COMMENT
Unpublished (1995)
Contact: Sverdllov ED
Structure and Function of Human Genes
Shemyakin Institute of Bioorganic Chemistry
16/10 Miklukho-Maklaya, Moscow, 117871, Russia
Tel: 70953306529
Fax: 70953306538
Email: sverdl@humgen.siobc.msk.su.
Location/Qualifiers
1..241
/db_xref="taxon:9606"

225777 241 bp mRNA linear EST 08-OCT-1998
pha-I-2R/1LW Outward Alu-primed hncDNA library Homo sapiens cDNA
clone pha-I-2R/1LW, mRNA sequence.

225777 241 bp mRNA linear EST 08-OCT-1998
pha-I-2R/1LW Outward Alu-primed hncDNA library Homo sapiens cDNA
clone pha-I-2R/1LW, mRNA sequence.

225777 241 bp mRNA linear EST 08-OCT-1998
pha-I-2R/1LW Outward Alu-primed hncDNA library Homo sapiens cDNA
clone pha-I-2R/1LW, mRNA sequence.

225777 241 bp mRNA linear EST 08-OCT-1998
pha-I-2R/1LW Outward Alu-primed hncDNA library Homo sapiens cDNA
clone pha-I-2R/1LW, mRNA sequence.

225777 241 bp mRNA linear EST 08-OCT-1998
pha-I-2R/1LW Outward Alu-primed hncDNA library Homo sapiens cDNA
clone pha-I-2R/1LW, mRNA sequence.

225777 241 bp mRNA linear EST 08-OCT-1998
pha-I-2R/1LW Outward Alu-primed hncDNA library Homo sapiens cDNA
clone pha-I-2R/1LW, mRNA sequence.

225777 241 bp mRNA linear EST 08-OCT-1998
pha-I-2R/1LW Outward Alu-primed hncDNA library Homo sapiens cDNA
clone pha-I-2R/1LW, mRNA sequence.

225777 241 bp mRNA linear EST 08-OCT-1998
pha-I-2R/1LW Outward Alu-primed hncDNA library Homo sapiens cDNA
clone pha-I-2R/1LW, mRNA sequence.

```

/clone="pha-i-2R/1LW"
/clone_lib="Outward Alu-primed hmcDNA library"
/clone="Vector: pGEM-32; Site_1: EcoRI; Site_2: BamHI: The
library was constructed as described in [Obradovic, D.,
Borodin, A.M., Kopantsev, E.P., Wagner, L.L., Volik, S.V.,
Ermolaeva, O.D., Lebedev, Y.B., Monastyrskaya, G.S.,
Sverdlov, E.D. (1993) Bioorganicheskaya khimiya, 20,
919-930]. This protocol is based on nested primer strategy
using Alu- specific primers (ALN3 and TC-65) that direct
the hmcDNA synthesis outward of Alu repeats."
BASE COUNT      72.a  64 c  44 g  55 t
ORIGIN

```

```

Query Match      0.9%; Score 39; DB 14; Length 241;
Best Local Similarity 100.0%; Pred. No. 4.1e-09;
Matches 39; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

```

QY  2 AGCTTGCATGCTCGAGTCGACTCTAGAGGATCCGGTT 40
      |||||||
Db   7 AGCTTGCATGCTCGAGTCGACTCTAGAGGATCCGGTT 45

```

```

RESULT 2
B60048/c
LOCUS      B60048      534 bp      DNA      linear      GSS 21-JUN-1998
DEFINITION CIT-HSP-385K18.TFB CIT-HSP Homo sapiens genomic clone 385K18, DNA
sequence.
ACCESSION  B60048
VERSION    B60048.1 GI:2614766
KEYWORDS   GSS.
SOURCE     human.
ORGANISM   Homo sapiens

```

```

REFERENCE
AUTHORS    Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
            1 (bases 1 to 534)
            Adams,M.D., Rounsley,S.D., Field,C.E., Bass,S., Linher,K., Golden
            ,K., Berry,K., Granger,D., Suh,E., Wible,C., Shizuya,H., Simon,M.
            and Venter,J.C.
            Use of a random BAC End Sequence Database for Sequence-Ready Map
            Building
            Unpublished (1997)
            Other_GSSs: CIT-HSP-385K18.TRB
            Contact: Mark Adams
            Department of Eukaryotic Genomics
            The Institute for Genomic Research
            9712 Medical Center Dr., Rockville, MD 20850, USA
            Tel: 301 838 0200
            Fax: 301 838 0208
            Email: mdadams@tigr.org
            Clones are available from Research Genetics (info@resgen.com). BAC
            end search page:
            http://www.tigr.org/tldb/humgen/bac_end_search/bac_end_search.html
            Seq primer: M13-21
            Class: BAC ends.

```

```

FEATURES
source      Location/Qualifiers
            1..534
            /organism="Homo sapiens"
            /db_xref="GDB:5378802"
            /db_xref="taxon:9606"
            /clone="385K18"
            /clone_lib="CIT-HSP"
            /sex="Male"
            /cell_type="Sperm"
            /note="Vector: pBelOBAC11; Site_1: HindIII; Site_2:
            HindIII"

```

```

BASE COUNT      175 a  145 c  102 g  112 t
ORIGIN
Query Match      0.9%; Score 38; DB 17; Length 534;
Best Local Similarity 100.0%; Pred. No. 1.3e-08;
Matches 38; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

```

QY  1 AAGCTTGATCGCTCGAGTCGACTCTAGAGGATCCGG 38
      |||||||
Db   47 AAGCTTGATCGCTCGAGTCGACTCTAGAGGATCCGG 10

```

```

RESULT 4
B70567/c
LOCUS      B70567      672 bp      DNA      linear      GSS 21-JUN-1998
DEFINITION CIT-HSP-2059L21.TF CIT-HSP Homo sapiens genomic clone 2059L21, DNA
sequence.
ACCESSION  B70567
VERSION    B70567.1 GI:2709791
KEYWORDS   GSS.
SOURCE     human.
ORGANISM   Homo sapiens

```

```

REFERENCE
AUTHORS    Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
            1 (bases 1 to 672)
            Adams,M.D., Rounsley,S.D., Field,C.E., Bass,S., Linher,K., Golden
            and Venter,J.C.
            Use of a random BAC End Sequence Database for Sequence-Ready Map
            Building
            Unpublished (1997)
            Other_GSSs: CIT-HSP-2063C19.TR
            Contact: Mark Adams
            Department of Eukaryotic Genomics
            The Institute for Genomic Research
            9712 Medical Center Dr., Rockville, MD 20850, USA
            Tel: 301 838 0200
            Fax: 301 838 0208
            Email: mdadams@tigr.org
            Clones are available from Research Genetics (info@resgen.com). BAC
            end search page:
            http://www.tigr.org/tldb/humgen/bac_end_search/bac_end_search.html
            Seq primer: M13-21
            Class: BAC ends.

```

```

BASE COUNT      176 a  147 c  139 g  155 t
ORIGIN
Query Match      0.9%; Score 38; DB 17; Length 617;
Best Local Similarity 100.0%; Pred. No. 1.2e-08;
Matches 38; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

```

QY  1 AAGCTTGATCGCTCGAGTCGACTCTAGAGGATCCGG 38
      |||||||
Db   47 AAGCTTGATCGCTCGAGTCGACTCTAGAGGATCCGG 10

```

```

RESULT 4
B70567/c
LOCUS      B70567      672 bp      DNA      linear      GSS 21-JUN-1998
DEFINITION CIT-HSP-2059L21.TF CIT-HSP Homo sapiens genomic clone 2059L21, DNA
sequence.
ACCESSION  B70567
VERSION    B70567.1 GI:2709791
KEYWORDS   GSS.
SOURCE     human.
ORGANISM   Homo sapiens

```

```

REFERENCE
AUTHORS    Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
            1 (bases 1 to 672)
            Adams,M.D., Rounsley,S.D., Field,C.E., Bass,S., Linher,K., Golden
            and Venter,J.C.
            Use of a random BAC End Sequence Database for Sequence-Ready Map
            Building
            Unpublished (1997)
            Other_GSSs: CIT-HSP-2063C19.TR
            Contact: Mark Adams
            Department of Eukaryotic Genomics
            The Institute for Genomic Research
            9712 Medical Center Dr., Rockville, MD 20850, USA
            Tel: 301 838 0200
            Fax: 301 838 0208
            Email: mdadams@tigr.org
            Clones are available from Research Genetics (info@resgen.com). BAC
            end search page:
            http://www.tigr.org/tldb/humgen/bac_end_search/bac_end_search.html
            Seq primer: M13-21
            Class: BAC ends.

```

```

BASE COUNT      176 a  147 c  139 g  155 t
ORIGIN
Query Match      0.9%; Score 38; DB 17; Length 617;
Best Local Similarity 100.0%; Pred. No. 1.2e-08;
Matches 38; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

```

QY  1 AAGCTTGATCGCTCGAGTCGACTCTAGAGGATCCGG 38
      |||||||
Db   47 AAGCTTGATCGCTCGAGTCGACTCTAGAGGATCCGG 10

```

K., Berry, K., Granger, D., Suh, E., Wible, C., Shizuya, H., Simon, M. and Venter, J.C.
 Use of a random BAC End Sequence Database for Sequence-Ready Map Building
 Unpublished (1997)
 Other_GSSs: CIT-HSP-2059L21.TR
 Contact: Mark Adams
 Department of Eukaryotic Genomics
 The Institute for Genomic Research
 9712 Medical Center Dr., Rockville, MD 20850, USA
 Tel: 301 838 0200
 Fax: 301 838 0208
 Email: mdadams@tigr.org
 Clones are available from Research Genetics (info@resgen.com). BAC end search page:
http://www.tigr.org/tdb/humgen/bac_end_search/bac_end_search.html
 Seq primer: M13-21
 Class: BAC ends.

FEATURES

source

Location/Qualifiers
 1..672
 /organism="Homo sapiens"
 /db_xref="GDB:7060128"
 /db_xref="taxon:9606"
 /clone="2059L21"
 /clone_lib="CIT-HSP"
 /sex="Male"
 /cell_type="Sperm"
 /note="Vector: pBelOBAC11; Site_1: HindIII; Site_2: HindIII"

BASE COUNT 195 a 118 c 118 g 241 t

Query Match 0.9%; Score 38; DB 17; Length 672;
 Best Local Similarity 100.0%; Pred. No. 1.2e-08;
 Matches 38; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 AAGCTTGATCGCTCAGGTGACACTCTAGAGATCCGG 38
 ||||||||||||||||||||||||||||||||||||||||
 Db 38 AAGCTTGATCGCTCAGGTGACACTCTAGAGATCCGG 1

RESULT 5

B64148/c

LOCUS B64148 739 bp DNA linear GSS 21-JUN-1998
 DEFINITION CIT-HSP-2021E8.TF CIT-HSP Homo sapiens genomic clone 2021E8, DNA sequence.

ACCESSION B64148

VERSION B64148.1 GI:2638138

KEYWORDS GSS.

SOURCE human.

ORGANISM

Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

REFERENCE

AUTHORS

1 (bases 1 to 739)
 Adams, M.D., Rounsley, S.D., Field, C.E., Bass, S., Linher, K., Golden, K., Berry, K., Granger, D., Suh, E., Wible, C., Shizuya, H., Simon, M. and Venter, J.C.

TITLE

Use of a random BAC End Sequence Database for Sequence-Ready Map Building

JOURNAL

COMMENT

Unpublished (1997)
 Contact: Mark Adams
 Department of Eukaryotic Genomics
 The Institute for Genomic Research
 9712 Medical Center Dr., Rockville, MD 20850, USA
 Tel: 301 838 0200
 Fax: 301 838 0208
 Email: mdadams@tigr.org
 Clones are available from Research Genetics (info@resgen.com). BAC end search page:
http://www.tigr.org/tdb/humgen/bac_end_search/bac_end_search.html
 Seq primer: M13-21
 Class: BAC ends.

FEATURES

Location/Qualifiers

source

1..739
 /organism="Homo sapiens"
 /db_xref="GDB:7045354"
 /db_xref="taxon:9606"
 /clone="2021E8"
 /clone_lib="CIT-HSP"
 /sex="Male"
 /cell_type="Sperm"
 /note="Vector: pBelOBAC11; Site_1: HindIII; Site_2: HindIII"

BASE COUNT 246 a 131 c 176 g 186 t

ORIGIN

Query Match 0.9%; Score 38; DB 17; Length 739;
 Best Local Similarity 100.0%; Pred. No. 1.2e-08;
 Matches 38; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 AAGCTTGATCGCTCAGGTGACACTCTAGAGATCCGG 38
 ||||||||||||||||||||||||||||||||||||||||
 Db 61 AAGCTTGATCGCTCAGGTGACACTCTAGAGATCCGG 24

RESULT 6

AQ916288/c

LOCUS

DEFINITION AQ916288 763 bp DNA linear GSS 02-DEC-1999
 clone nb06063017f CUGI Rice BAC Library (EcoRI) Oryza sativa genomic clone nb06063017f, DNA sequence.

ACCESSION AQ916288

VERSION AQ916288.1 GI:6512804

KEYWORDS GSS.

SOURCE

ORGANISM

Oryza sativa
 Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
 Ehrhartoideae; Oryzoae; Oryza.

REFERENCE

AUTHORS

TITLE

JOURNAL

COMMENT

1 (bases 1 to 763)
 Wing, R.A. and Dean, R.A.
 A BAC End Sequencing Framework to Sequence the Rice Genome
 Unpublished (1998)
 Contact: Wing RA
 Clemson University Genomics Institute
 Clemson University
 100 Jordan Hall, Clemson, SC 29634, USA
 Tel: 864 656 7288
 Fax: 864 656 4293
 Email: rwing@clemson.edu

Seq primer: TAATAGGACTCACTATAGGG

Class: BAC ends

High quality sequence start: 43

High quality sequence stop: 412.

Location/Qualifiers

1..763

/organism="Oryza sativa"

/strain="Japonica"

/cultivar="Nipponbare"

/db_xref="taxon:4530"

/clone="nb06063017f"

/clone_lib="CUGI Rice BAC Library (EcoRI)"

/tissue_type="Leaf"

/lab_host="E. coli DH10B"

/note="Vector: pBACindigo; Site_1: EcoRI; Site_2: EcoRI;

Rice is the most important food crop in the world. Half of

the world population, especially those inhabiting highly

populated areas of the humid tropics and subtropics, rely

on rice as their primary source of carbohydrate.

Monocotyledonous rice is a diploid plant (2n=24) with a

haploid genome equivalent of 431 Mbp (Arumuganathan and

Earle, 1991). The relatively small genome of rice, three

times larger than that of Arabidopsis, makes it suitable

for genomic studies. In order to facilitate positional

cloning, physical mapping and genome sequencing of rice,

we have constructed a BAC library from Oryza sativa,

Nipponbare variety using EcoRI as the cloning enzyme. The

library contains 55,296 clones with an average insert size of 121 kb providing approximately 15 haploid genome equivalents. The deep coverage allows the isolation a particular sequence with a probability of 99.9 %. Three high density filters, each containing 18,432 clones (doubly spotted), represent the whole library for colony screening and can be requested from the Clemson University BAC/EST Resource Center (www.genome.clemson.edu)."

BASE COUNT 204 a 195 c 163 g 197 t 4 others
ORIGIN

Query Match 0.9%; Score 38; DB 17; Length 763;
Best Local Similarity 100.0%; Pred. No. 1.2e-08;
Matches 38; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy 1 AAGCTTCATCGCTCGAGTCGACTCTAGAGGATCCG 38
|||||
Db 68 AAGCTTGATCGCTCGAGTCGACTCTAGAGGATCCG 31
|||||

RESULT 7
B32178/c
LOCUS B32178 63 bp DNA linear GSS 17-OCT-1997
DEFINITION HS-1014-B2-E05-MF.abi CIT Human Genomic Sperm Library C Homo sapiens genomic clone Plate=CT 789 Col=10 Row=J, DNA sequence.
ACCESSION B32178
VERSION B32178.1 GI:2531547
KEYWORDS GSS
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

REFERENCE 1 (bases 1 to 63)
AUTHORS Mahairas G.G., Zackrone, K.D., Smith, T., Tipton, S., Schmidt, S., Traicoff, R., Abajian, C., Blanchard, A., West, A. and Hood, L.E.
TITLE Construction of a Characterized Clone Resource for Genomic Sequencing: Generation and Preliminary Analysis of 20,000 Sequence Tagged Connectors
JOURNAL Unpublished (1997)
COMMENT Contact: Mahairas GG, Zackrone KD, Hood L
University of Washington
Seattle, WA 98195, USA
Tel: (206) 616-8744
Fax: (206) 685-7301
Email: kzackrone@u.washington.edu
Sequence Tagged Connector
Plate: CT 789 row: J column: 10
Class: BAC ends
High quality sequence stop: 63.
Location/Qualifiers
1. .63
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="Plate=CT 789 Col=10 Row=J"
/clone_lib="CIT Human Genomic Sperm Library C"
/sex="M"
/note="Organ: sperm; Vector: pBelOBAC11; BAC Clones in E-Coli DH10B"

BASE COUNT 14 a 18 c 18 g 13 t
ORIGIN

Query Match 0.9%; Score 37; DB 17; Length 63;
Best Local Similarity 100.0%; Pred. No. 6.1e-08;
Matches 37; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy 1 AAGCTTCATCGCTCGAGTCGACTCTAGAGGATCCG 37
|||||
Db 51 AAGCTTGATCGCTCGAGTCGACTCTAGAGGATCCG 15
|||||

RESULT 8
B40882/c
LOCUS B40882 123 bp DNA linear GSS 18-OCT-1997

DEFINITION HS-1052-B1-F02-MF.abi CIT Human Genomic Sperm Library C Homo sapiens genomic clone Plate=CT 774 Col=3 Row=L, DNA sequence.
ACCESSION B40882
VERSION B40882.1 GI:2545134
KEYWORDS GSS.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

REFERENCE 1 (bases 1 to 123)
AUTHORS Mahairas G.G., Zackrone, K.D., Smith, T., Tipton, S., Schmidt, S., Traicoff, R., Abajian, C., Blanchard, A., West, A. and Hood, L.E.
TITLE Construction of a Characterized Clone Resource for Genomic Sequencing: Generation and Preliminary Analysis of 20,000 Sequence Tagged Connectors
JOURNAL Unpublished (1997)
COMMENT Contact: Mahairas GG, Zackrone KD, Hood L
University of Washington
Seattle, WA 98195, USA
Tel: (206) 616-8744
Fax: (206) 685-7301
Email: kzackrone@u.washington.edu
Sequence Tagged Connector
Plate: CT 774 row: L column: 3
Class: BAC ends
High quality sequence stop: 123.
Location/Qualifiers
1. .123
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="Plate=CT 774 Col=3 Row=L"
/clone_lib="CIT Human Genomic Sperm Library C"
/sex="M"
/note="Organ: sperm; Vector: pBelOBAC11; BAC Clones in E-Coli DH10B"

BASE COUNT 39 a 20 c 24 g 40 t
ORIGIN

Query Match 0.9%; Score 37; DB 17; Length 123;
Best Local Similarity 100.0%; Pred. No. 5.5e-08;
Matches 37; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy 1 AAGCTTCATCGCTCGAGTCGACTCTAGAGGATCCG 37
|||||
Db 49 AAGCTTGATCGCTCGAGTCGACTCTAGAGGATCCG 13
|||||

RESULT 9
B79682/c
LOCUS B79682 161 bp DNA linear GSS 24-OCT-1998
DEFINITION CIT-HSP-2045K8.TF CIT-HSP Homo sapiens genomic clone 2045K8, DNA sequence.
ACCESSION B79682
VERSION B79682.1 GI:2866705
KEYWORDS GSS.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

REFERENCE 1 (bases 1 to 161)
AUTHORS Adams, M.D., Rounsley, S.D., Zhao, S., Field, C.E., Bass, S., Linher, K., Golden, K., Berry, K., Granger, D., Suh, E., Wible, C., Shizuya, H., Simon, M. and Venter, J.C.
TITLE Use of a random BAC End Sequence Database for Sequence-Ready Map Building (1998)
JOURNAL Unpublished (1998)
COMMENT Contact: Mark Adams
Department of Eukaryotic Genomics
The Institute for Genomic Research
9712 Medical Center Dr., Rockville, MD 20850, USA
Tel: 301 838 0200
Fax: 301 838 0208
Email: mdadams@tigr.org

BASE COUNT 14 a 18 c 18 g 13 t
ORIGIN

Query Match 0.9%; Score 37; DB 17; Length 63;
Best Local Similarity 100.0%; Pred. No. 6.1e-08;
Matches 37; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy 1 AAGCTTCATCGCTCGAGTCGACTCTAGAGGATCCG 37
|||||
Db 51 AAGCTTGATCGCTCGAGTCGACTCTAGAGGATCCG 15
|||||

RESULT 8
B40882/c
LOCUS B40882 123 bp DNA linear GSS 18-OCT-1997

Clones are available from Research Genetics (info@resgen.com). BAC end search page:
http://www.tigr.org/tldb/hungen/bac_end_search/bac_end_search.html
 Seq primer: M13-21
 Class: BAC ends.

FEATURES

Location/Qualifiers

```
1..161
/organism="Homo sapiens"
/db_xref="GB:7054715"
/db_xref="taxon:9606"
/clone="2045K8"
/clone_lib="CIR-HSP"
/sex="Male"
/cell_type="Sperm"
/Note="Vector: pBelOBAC11; Site_1: HindIII; Site_2: HindIII"
```

BASE COUNT 47 a 42 c 37 g 35 t
 ORIGIN

Query Match 0.9%; Score 37; DB 17; Length 161;
 Best Local Similarity 100.0%; Pred. No. 5.2e-08;
 Matches 37; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 AAGCTTGATCGCTCGAGTGCAGTCTAGAGATCCG 37
 |||||||||||||||||||||||||||||||||||||||
 DB 59 AAGCTTGATCGCTCGAGTGCAGTCTAGAGATCCG 23

RESULT 10

BM851683/c
 LOCUS K-EST0132614 S21SNU520 Homo sapiens cDNA clone S21SNU520-42-B01 5',
 DEFINITION mRNA sequence.

ACCESSION BM851683
 VERSION BM851683.1 GI:19208082
 KEYWORDS EST.
 SOURCE human.

ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

REFERENCE 1 (bases 1 to 205)
 AUTHORS Kim,N.S., Hahn,Y., Oh,J.H., Lee,J.Y., Ahn,H.Y., Chu,M.Y., Kim,M.R.,
 Oh,K.J., Cheong,J.E., Sohn,H.Y., Kim,J.M., Park,H.S., Kim,S. and
 Kim,Y.S.

TITLE 21C Frontier Korean EST Project 2001

JOURNAL Unpublished (2002)

COMMENT Contact: Kim YS

Genome Research Center
 Korea Research Institute of Bioscience & Biotechnology
 52 Eoeun-dong Yuseong-gu, Daejeon 305-333, South Korea
 Tel: +82-42-860-4470
 Fax: +82-42-860-4409
 Email: yongsung@mail.kribb.re.kr

Plate: 42 row: B column: 01

High quality sequence stop: 205.

Location/Qualifiers

FEATURES

source

```
1..205
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="S21SNU520-42-B01"
/clone_lib="S21SNU520"
/sex="F"
/tissue_type="Stomach"
/cell_type="Floating aggregates"
/lab_host="Top10F"
```

/Note="Organ: Stomach; Vector: pTZ18Rpl; Site_1: EcoRI;
 Site_2: NotI; The poly (A)+ RNA was dephosphorylated with
 bacterial alkaline phosphatase (BAP) and then decapped
 with tobacco acid pyrophosphatase (TAP). The decapped
 intact mRNA was ligated with DNA-RNA linker including EcoR
 I site by treatment of T4 RNA ligase and the first strand
 cDNA was synthesized from oligo dt-selected mRNA by

priming with dt-tailed vector. The dt-tailed vector was
 adjusted to have about 60nt. The cDNA vector was
 circularized with E. coli DNA ligase after digestion of
 EcoRI which site is also included in vector. An RNA strand
 converted to a DNA strand by Okayama-Berg method. The
 obtained cDNA vectors were used for transformation of
 competent cells E. coli Top10F by electroporation method.
 The cDNA libraries constructed by this method are
 full-length enriched cDNA library."

BASE COUNT 39 a 59 c 63 g 44 t
 ORIGIN

Query Match 0.9%; Score 37; DB 14; Length 205;
 Best Local Similarity 100.0%; Pred. No. 5.1e-08;
 Matches 37; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 AAGCTTGATCGCTCGAGTGCAGTCTAGAGATCCG 37
 |||||||||||||||||||||||||||||||||||||||
 DB 174 AAGCTTGATCGCTCGAGTGCAGTCTAGAGATCCG 138

RESULT 11

BM850221/c

LOCUS

DEFINITION

ACCESSION

VERSION

KEYWORDS

SOURCE

ORGANISM

REFERENCE

AUTHORS

TITLE

JOURNAL

COMMENT

21C Frontier Korean EST Project 2001

Unpublished (2002)

Contact: Kim YS

Genome Research Center

Korea Research Institute of Bioscience & Biotechnology
 52 Eoeun-dong Yuseong-gu, Daejeon 305-333, South Korea

Tel: +82-42-860-4470

Fax: +82-42-860-4409

Email: yongsung@mail.kribb.re.kr

Plate: 41 row: A column: 10

High quality sequence stop: 206.

Location/Qualifiers

FEATURES

source

```
1..206
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="S21SNU520-41-A10"
/clone_lib="S21SNU520"
/sex="F"
/tissue_type="Stomach"
/cell_type="Floating aggregates"
/lab_host="Top10F"
```

/Note="Organ: Stomach; Vector: pTZ18Rpl; Site_1: EcoRI;
 Site_2: NotI; The poly (A)+ RNA was dephosphorylated with
 bacterial alkaline phosphatase (BAP) and then decapped
 with tobacco acid pyrophosphatase (TAP). The decapped
 intact mRNA was ligated with DNA-RNA linker including EcoR
 I site by treatment of T4 RNA ligase and the first strand
 cDNA was synthesized from oligo dt-selected mRNA by
 priming with dt-tailed vector. The dt-tailed vector was
 adjusted to have about 60nt. The cDNA vector was
 circularized with E. coli DNA ligase after digestion of
 EcoRI which site is also included in vector. An RNA strand
 converted to a DNA strand by Okayama-Berg method. The
 obtained cDNA vectors were used for transformation of
 competent cells E. coli Top10F by electroporation method.

The cDNA libraries constructed by this method are full-length enriched cDNA library."

39 a 56 c 65 g 46 t

Query Match 0.9%; Score 37; DB 14; Length 206;
Best Local Similarity 100.0%; Pred. No. 5.1e-08;
Matches 37; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 AAGCTTGATCGCTGCAGGTGACTCTAGAGGATCCG 37
|||||
Db 170 AAGCTTGATCGCTGCAGGTGACTCTAGAGGATCCG 134
|||||

RESULT 12
CNS00271/c
LOCUS
DEFINITION Drosophila melanogaster genome survey sequence T7 end of BAC
BACN01A13 of DrosBAC library from Drosophila melanogaster (fruit fly), genomic survey sequence.
ACCESSION AL097303.1 GI:5608914
VERSION
KEYWORDS
SOURCE Drosophila melanogaster.
ORGANISM Drosophila melanogaster.
Eukaryota; Metazoa; Arthropoda; Insecta; Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.
REFERENCE 1 (bases 1 to 207)
AUTHORS
TITLE Direct Submission
JOURNAL Submitted (23-JUL-1999) Genoscope - Centre National de Sequencage : BP 191 91006 EVRY cedex - FRANCE (E-mail : seqref@genoscope.cns.fr - Web : www.genoscope.cns.fr)
COMMENT Determination of this BAC-end sequence was carried out as part of a collaboration with the European Drosophila Genome Project (EDGP) - http://www.edgp.ebi.ac.uk -. This Drosophila melanogaster BAC library (Dros BAC) was made by Alain Billaud at CEPH (Centre d'Etude du Polymorphisme Humain) with funding provided by a MRC project grant. The DNA was prepared from embryos by Alain Bucheton and Genevieve Payan. It has been constructed in the vector pBelobAC11.

FEATURES
source
1..207
Location/Qualifiers
/organism="Drosophila melanogaster"
/db_xref="taxon:7227"
/clone="BACN01A13"
/clone_lib="DrosBAC"
/plasmid="pBelobAC11"
/note="end : T7"

BASE COUNT 40 a 62 c 40 g 62 t 3 others
ORIGIN

Query Match 0.9%; Score 37; DB 17; Length 207;
Best Local Similarity 100.0%; Pred. No. 5e-08;
Matches 37; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 AAGCTTGATCGCTGCAGGTGACTCTAGAGGATCCG 37
|||||
Db 187 AAGCTTGATCGCTGCAGGTGACTCTAGAGGATCCG 151
|||||

RESULT 13
B33347/c
LOCUS
DEFINITION HS-1017-A2-B08-MF.abi CIT Human Genomic Sperm Library C Homo sapiens genomic clone Plate=CT 792 Col=16 Row=C, DNA sequence.
ACCESSION B33347
VERSION B33347.1 GI:2532716
KEYWORDS GSS.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

Mammalia; Eutheria; Primates; Catarrhini; Hominoidea; Homo.
1 (bases 1 to 210)
Mahairas,G.G., Zackrone,K.D., Smith,T., Tipton,S., Schmidt,S., Traicoff,R., Abajian,C., Blanchard,A., West,A. and Hood,L.E.
Construction of a Characterized Clone Resource for Genomic Sequencing: Generation and Preliminary Analysis of 20,000 Sequence Tagged Connectors
Unpublished (1997)
Contact: Mahairas GG, Zackrone KD, Hood L
University of Washington
Seattle, WA 98195, USA
Tel: (206) 616-8744
Fax: (206) 685-7301
Email: kzackron@u.washington.edu
Sequence Tagged Connector
Plate: CT 792 row: C column: 16
Class: BAC ends
High quality sequence stop: 210.
Location/Qualifiers
1..210
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="Plate=CT 792 Col=16 Row=C"
/clone_lib="CIT Human Genomic Sperm Library C"
/sex="M"
/note="Organ: sperm; Vector: pBelobAC11; BAC Clones in E-Coli DH10B"

BASE COUNT 53 a 48 c 44 g 65 t
ORIGIN

Query Match 0.9%; Score 37; DB 17; Length 210;
Best Local Similarity 100.0%; Pred. No. 5e-08;
Matches 37; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 AAGCTTGATCGCTGCAGGTGACTCTAGAGGATCCG 37
|||||
Db 51 AAGCTTGATCGCTGCAGGTGACTCTAGAGGATCCG 15
|||||

RESULT 14
AT000691/c
LOCUS
DEFINITION AT000691 Brassica rapa guard cell Brassica rapa subsp. pekinensis
cDNA clone DGT694, mRNA sequence.
ACCESSION AT000691
VERSION AT000691.1 GI:3414225
KEYWORDS EST.
SOURCE Brassica rapa subsp. pekinensis.
ORGANISM Brassica rapa subsp. pekinensis.
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae; eurosids II; Brassicales; Brassicaceae; Brassica.
REFERENCE 1 (bases 1 to 218)
AUTHORS Kwak,J.M., Kim,S.A., Hong,S.W. and Nam,H.G.
TITLE Evaluation of 515 expressed sequence tags obtained from guard cells of Brassica campestris
JOURNAL Planta 202 (1), 9-17 (1997)
MEDLINE 97320163
COMMENT Contact: Hong-Gil Nam
Department of Life Science, Plant Molecular Genetics Laboratory
Pohang University of Science and Technology
San 31 Hyojadong, Pohang Kyungbuk 790-784, Korea
Email: hgn@bric.postech.ac.kr
Submitted through BRIC(Biological Research Information Center) of Korea URL: http://bric.postech.ac.kr/.

FEATURES
source
1..218
Location/Qualifiers
/organism="Brassica rapa subsp. pekinensis"
/db_xref="taxon:51351"
/clone="DGT694"
/clone_lib="Brassica rapa guard cell"
/cell_type="guard cell protoplast"

BASE COUNT 60 a 42 c 29 g 87 t

ORIGIN

Query Match 0.9%; Score 37; DB 9; Length 218;
Best Local Similarity 100.0%; Pred. No. 5e-08;
Matches 37; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 AAGCTTGATCGCTGCAGGTGACTCTAGAGGATCCG 37
|||||
Db 217 AAGCTTGATCGCTGCAGGTGACTCTAGAGGATCCG 181
|||||

RESULT 15

AQ006733/c
LOCUS
DEFINITION CIT-HSP-2291D16.TF CIT-HSP Homo sapiens genomic clone 2291D16, DNA
221 bp DNA linear GSS 27-JUN-1998
sequence.
ACCESSION AQ006733
VERSION AQ006733.1 GI:3083959
KEYWORDS GSS.
SOURCE human.

ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

REFERENCE 1 (bases 1 to 221)
Adams,M.D., Rounsley,S.D., Zhao,S., Field,C.E., Bass,S., Linher,K.,
Golden,K., Berry,K., Granger,D., Suh,E., Wible,C., Shizuya,H.,
Simon,M. and Venter,J.C.
Use of a random BAC End Sequence Database for Sequence-Ready Map
Building (1998)

JOURNAL Unpublished (1998)
COMMENT Other_GSSs: CIT-HSP-2291D16.FR
Contact: Mark Adams
Department of Eukaryotic Genomics
The Institute for Genomic Research
9712 Medical Center Dr., Rockville, MD 20850, USA
Tel: 301 838 0200
Fax: 301 838 0208
Email: mdadams@tigr.org
Clones are available from Research Genetics (info@resgen.com). BAC
end search page:
http://www.tigr.org/tldb/hungen/bac_end_search/bac_end_search.html
Seq primer: M13-21;
Class: BAC ends.

FEATURES

source Location/Qualifiers
1..221
/organism="Homo sapiens"
/db_xref="GDB:7150963"
/db_xref="taxon:9606"
/clone="2291D16"
/clone_lib="CIT-HSP"
/sex="Male"
/cell_type="Sperm"
/note="Vector: pBelobAC11; Site_1: HindIII; Site_2:
HindIII"

BASE COUNT 48 a 63 c 48 g 62 t
ORIGIN

Query Match 0.9%; Score 37; DB 17; Length 221;
Best Local Similarity 100.0%; Pred. No. 5e-08;
Matches 37; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 AAGCTTGATCGCTGCAGGTGACTCTAGAGGATCCG 37
|||||
Db 77 AAGCTTGATCGCTGCAGGTGACTCTAGAGGATCCG 41
|||||

Search completed: May 19, 2003, 09:35:37
Job time : 5405 secs

THIS PAGE BLANK (U&PTO)

THIS PAGE BLANK (U&PTO)

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: May 18, 2003, 18:00:11 ; Search time 10356 Seconds
(without alignments)
11564.122 Million cell updates/sec

Title: US-09-926-163B-1
Perfect score: 4115
Sequence: 1 aagctgtcatgcctgcaggt.....tccggcgccgctgaagctt 4115

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 2054640 seqs, 14551402878 residues

Total number of hits satisfying chosen parameters: 4109280

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : GenEmbl.*

- 1: gb_ba.*
- 2: gb_htg.*
- 3: gb_in.*
- 4: gb_on.*
- 5: gb_ov.*
- 6: gb_pet.*
- 7: gb_ph.*
- 8: gb_pl.*
- 9: gb_pr.*
- 10: gb_ro.*
- 11: gb_sts.*
- 12: gb_sy.*
- 13: gb_un.*
- 14: gb_vi.*
- 15: em_ba.*
- 16: em_fun.*
- 17: em_hum.*
- 18: em_in.*
- 19: em_mu.*
- 20: em_on.*
- 21: em_or.*
- 22: em_ov.*
- 23: em_pat.*
- 24: em_ph.*
- 25: em_pl.*
- 26: em_ro.*
- 27: em_sts.*
- 28: em_un.*
- 29: em_vi.*
- 30: em_htg_hum.*
- 31: em_htg_inv.*
- 32: em_htg_other.*
- 33: em_htg_mus.*
- 34: em_htg_pln.*
- 35: em_htg_rod.*
- 36: em_htg_mam.*
- 37: em_htg_virt.*
- 38: em_sy.*
- 39: em_htgo_hum.*
- 40: em_htgo_mus.*
- 41: em_htgo_other.*

Pred. No. is the number of results predicted by chance to have a

score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match %	Length	DB	ID	Description
1	4073	99.0	4073	1	AB028937	AB028937 Gluconoba
2	955.6	23.2	1458	1	AB063188	AB063188 Gluconoba
3	955.6	23.2	1458	6	AX012443	AX012443 Sequence
4	955.6	23.2	1458	6	E50685	E50685 Gene manipu
5	351.4	8.5	10977	1	AE004660	AE004660 Pseudomon
6	327.4	8.0	10031	1	AF007800	AF007800 Pseudomon
7	292.4	7.1	332635	1	AP003005	AP003005 Mesorhizo
8	268.4	6.5	323450	1	SME591790	AL591790 Sinorhizo
9	262.2	6.4	189370	1	AF010496	AF010496 Rhodobact
10	259.2	6.3	11164	1	AE009374	AE009374 Agrobacte
11	259.2	6.3	14945	1	AE008240	AE008240 Agrobacte
12	247.4	6.0	9810	1	AF018073	AF018073 Rhodobact
13	236.8	5.8	10955	1	AE005663	AE005663 Escherich
14	236.8	5.8	168764	1	AP002569	AP002569 Escherich
15	236.2	5.7	189333	1	RME603647	AL603647 Rhizobium
16	224	5.4	4539	1	DI3329	D13329 Escherichia
17	224	5.4	11277	1	AE000503	AE000503 Escherich
18	224	5.4	338534	1	ECOUW93	U14003 Escherichia
19	219.4	5.3	11047	1	AE012535	AE012535 Xanthomon
20	213.6	5.2	1509	6	AX120210	AX120210 Sequence
21	213.6	5.2	1632	6	AX065209	AX065209 Sequence
22	213.6	5.2	340000	1	AP005274	AP005274 Corynebac
23	213.6	5.2	349980	6	AX120085	AX120085 Sequence
24	210.2	5.1	10788	1	AE012076	AE012076 Xanthomon
25	210	5.1	266050	1	AL627270	AL627270 Salmonell
26	206.8	5.0	21585	1	AE008765	AE008765 Salmonell
27	203.2	4.9	25034	1	AE008844	AE008844 Salmonell
28	202.8	4.9	10594	1	AE005823	AE005823 Caulobact
29	198.4	4.8	230050	1	AL627277	AL627277 Salmonell
30	186.8	4.5	14681	1	AE013893	AE013893 Yersinia
31	186.8	4.5	199050	1	AJ414147	AJ414147 Yersinia
32	184	4.5	24336	1	AF008841	AF008841 Salmonell
33	181.8	4.4	347750	1	AP002998	AP002998 Mesorhizo
34	179	4.3	286485	1	AP002557	AP002557 Escherich
35	178	4.3	11783	1	AE005350	AE005350 Escherich
36	177.4	4.3	11258	1	AE000251	AE000251 Escherich
37	177.4	4.3	14045	1	D90797	D90797 E.coli geno
38	177.4	4.3	19465	1	D90798	D90798 E.coli geno
39	173.4	4.2	10446	1	AE000306	AE000306 Escherich
40	173.4	4.2	10449	1	AE005449	AE005449 Escherich
41	173.4	4.2	75888	1	ECOHU47	U00007 47 to 48 ce
42	173.4	4.2	268857	1	AP002560	AP002560 Escherich
43	171.8	4.2	17093	1	D90849	D90849 E.coli geno
44	170.6	4.1	244894	2	AC084324	AC084324 Mus muscu
45	167	4.1	3250	8	SCYNR073C	Z71688 S.cerevisia

ALIGNMENTS

RESULT 1	AB028937	4073 bp	DNA	linear	BCT 11-JUL-2000
LOCUS	Gluconobacter oxydans gene for sorbitol dehydrogenase, complete cds.				
DEFINITION	Gluconobacter oxydans gene for sorbitol dehydrogenase, complete cds.				
ACCESSION	AB028937				
VERSION	AB028937.1 GI:9049448				
KEYWORDS	sorbitol dehydrogenase.				
SOURCE	Gluconobacter oxydans (strain:G624) DNA.				
ORGANISM	Gluconobacter oxydans				
REFERENCE	Bacteria: Proteobacteria; alpha subdivision; Acetobacteraceae; Gluconobacter.				
AUTHORS	1 (sites) Shibata,T., Ichikawa,C., Matsuura,M., Takata,Y., Noguchi,Y., Saito,Y. and Yamashita,M.				

TITLE Cloning of a gene for D-Sorbitol dehydrogenase from *Gluconobacter*
oxydans G624 and expression of the gene in *Pseudomonas putida*
IFO3738
J. Biosci. Bioeng. 89, 463-468 (2000)
2. (bases 1 to 4073)
Shibata, T. and Saito, Y.
REFERENCE Direct Submission
TITLE Submitted (15-JUN-1999) Takashi Shibata, Fujisawa Pharmaceutical
Co., Ltd., Fermentation Development Laboratories, Shinkawa-cho,
Nakagawara 156, Nishikasuga-gun, Aichi 452-0915, Japan
(E-mail: takashi.shibata@fujisawa.co.jp, Tel: 81-52-401-0275,
Fax: 81-52-400-1380)
FEATURES
Location/Qualifiers
1. 4073
/organism="Gluconobacter oxydans"
/strain="G624"
/db_xref="taxon:442"
501..1958
CDS
/codon_start=1
/transl_table=11
/product="sorbitol dehydrogenase"
/protein_id="BAA99414.1"
/db_xref="GI:904949"
/translation="MITRETLKSLPANVQAPPYDIDGKIVHFGVGNFFRAHEAFY
VEQLEHAPDWAIWVGVLGSDRSKKAEEFKADQCLYSLTEAPSGSKTVRVNGALR
DYLLAPADPEAVLWLPDPAIRIVSMITTEGGYINNETGAFDLENAARADLNKPK
PSTVFGVVEALRRWDAGKAFVMSCDNLNRHNGVARKAFGLYAKARPELAKWIE
ENATFPNGVDRIPTVSAEIAKLNAAAGLDDPLVAEDFHOWVLEDFADGRPEL
GQWQVMDVDWEYKIRMLNAGHVMCLFPGLVGTENVDDAIEDSELGNLKNLYL
NKDVIPLKAPSGMTLEGYRDSVIRFSNKAUSDOTLRIASDGCSKVQVFWETVRR
IEDRDLRSIAFGFIASLEMLRGGGTVESEPTGYGDAEMKLAKADDFESSLKL
AFDGRDLDTSELQKIVLIRKREKGVAAIPA"
BASE COUNT 823 a 1122 c 1259 g 869 t
ORIGIN
Query Match 99.0%; Score 4073; DB 1; Length 4073;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 4073; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 37 GGTTTTGGCAGCGTCCCTAGATTGATCGCGCGTCTGTTGACCGACATGATGCTGGTGGC 96
DB 1 GGTTTTGGCAGCGTCCCTAGATTGATGCGCGCTCTGTTGACCGACATGATGCTGGTGGC 60
QY 97 ACGTGCCATTGCGACGGGGCGTGGACCGGGAACACAGCGCTGCTGCTTTGTACAAGG 156
DB 61 ACGTGCCATTGCGACGGGGCGTGGACCGGGAACACAGCGCTGCTGCTTTGTACAAGG 120
QY 157 GCTGAGTCATGCGTGGTGGTGGCGACATAGTTGGGAAGACAGTTGCGCGCAAGCA 216
DB 121 GCTGAGTCATGCGTGGTGGTGGCGACATAGTTGGGAAGACAGTTGCGCGCAAGCA 180
QY 217 GAACACAGCATGAACAGCAGTCCGAAGACAGGAATCTCGGCTCTCTACCGGATTTGA 276
DB 181 GAACACAGCATGAACAGCAGTCCGAAGACAGGAATCTCTCGGCTCTCTACCGGATTTGA 240
QY 277 AGACAGACCGCTCTGAGATGCGTTTGTGATGTCCTGTTCCGCGAGGATCTCGAACG 336
DB 241 AGACAGACCGCTCTGAGATGCGTTTGTGATGTCCTGTTCCGCGAGGATCTCGAACG 300
QY 337 GCGTGTGGGTGCTCATGCGTCTGATGCGAGTGGCGCAAAAGGTCTCTGAACAGACG 396
DB 301 GCGTGTGGGTGCTCATGCGTCTGATGCGAGTGGCGCAAAAGGTCTCTGAACAGACG 360
QY 397 TCCCGCGAGGCGAGTCAAGGTGCAATGCTCTGTTGAAACCGTCAATTCGGTTTGA 456
DB 361 TCCCGCGAGGCGAGTCAAGGTGCAATGCTCTGTTGAAACCGTCAATTCGGTTTGA 420
QY 457 CGTTGTTTCGGGGTATGATGCGACATGCCGGGCTTGTGCGTCCCGCTCAGCGACCGG 516
DB 421 CGTTGTTTCGGGGTATGATGCGACATGCCGGGCTTGTGCGTCCCGCTCAGCGACCGG 480
QY 517 CCGAACCACGAGATTCATGATTAGCGCGCAACCCCTTAAGTCTCTTCTCCCAATG 576
DB 15361 GGGACACGCGTATCAGCGCGTTTCTCCCAACAGGCGATGTCTGGACGACGCGTCCGCGATTG 1620

481 CCGAACCACGAGAGATTCCATGATTACGGCGCAAAACCTTAAGTCTCTTCTCCGCAATG 540
577 TCCAGGCTCCCGCTATGACATCGACGGATCAAGCTGGATCGTCAATTCGCTAG 636
541 TCCAGGCTCCCGCTATGACATCGACGGATCAAGCTGGATCGTCAATTCGCTAG 600
637 GTAACCTTTTTCGAGCGCCATGAGCGTTCACGTCGAGCAGATTCCTTGAACACGCTCCGG 696
601 GTAACCTTTTTCGAGCGCCATGAGCGTTCACGTCGAGCAGATTCCTTGAACACGCTCCGG 660
697 ACTGGCGGATTTGGTGGTGGCTGACGGCGAGTACCGTTTCAAGAAAAAAGCCGAGG 756
661 ACTGGCGGATTTGGTGGTGGCTGACGGCGAGTACCGTTTCAAGAAAAAAGCCGAGG 720
757 AATTCAAGGCCCGAGGACTGCTGTTATTCCTGACCGAGAGCGCTCCGTCGGCAAGAGCA 816
721 AATTCAAGGCCCGAGGACTGCTGTTATTCCTGACCGAGAGCGCTCCGTCGGCAAGAGCA 780
817 CGGTGCGCGTATGCGCGCGCTGCTGACTATCTGCTTGGCCCGCGCGATCCGGAAGCGG 876
781 CGGTGCGCGTATGCGCGCGCTGCTGACTATCTGCTTGGCCCGCGCGATCCGGAAGCGG 840
877 TGCTGAAGCATCTTTGTTGATCCCGGCCATCCGCATCTGCTTCCATGACGATCAGCAAGCGG 936
841 TGCTGAAGCATCTTTGTTGATCCCGGCCATCCGCATCTGCTTCCATGACGATCAGCAAGCGG 900
937 GCTAACAATCAACGACGACGCGTGGTTCGATCTGGAGATGCGGCAAGTAAGCGG 996
901 GCTAACAATCAACGACGACGCGTGGTTCGATCTGGAGATGCGGCAAGTAAGCGG 960
997 ACCTCAAGAACCCCGAAAAACCGCTCTACCGCTTTTCGGTTACGTGGTGGAGGCCCTGGGTC 1056
961 ACCTCAAGAACCCCGAAAAACCGCTCTACCGCTTTTCGGTTACGTGGTGGAGGCCCTGGGTC 1020
1057 GTCGTTGGGATGCGCGTGGTAAGCATTTACGCTCATGCTCTGTGATAACCTGCGTCATA 1116
1021 GTCGTTGGGATGCGCGTGGTAAGCATTTACGCTCATGCTCTGTGATAACCTGCGTCATA 1080
1117 ACGGCAATGTCGCGCGCAAGCGCTTCTCGGCTATGCGAAGCGCGGATCCGCGATTGG 1176
1081 ACGGCAATGTCGCGCGCAAGCGCTTCTCGGCTATGCGAAGCGCGGATCCGCGATTGG 1140
1177 CGAAGTGGATTGAGAAAAACGCGACCTTCCCGAAGCAAGTGGTGGATCGCATACCCCGA 1236
1141 CGAAGTGGATTGAGAAAAACGCGACCTTCCCGAAGCAAGTGGTGGATCGCATACCCCGA 1200
1237 CCGTTTCGCGGGAATCGCAAGAGCTCAACCGCGCGAGTGGCTGGATGACGACCTGC 1296
1201 CCGTTTCGCGGGAATCGCAAGAGCTCAACCGCGCGAGTGGCTGGATGACGACCTGC 1260
1297 CCGTTTCGCGGAGATTTCATCAGTGGTGGTGGAGACACAGTTTTCGCGATGGCGTC 1356
1261 CCGTTTCGCGGAGATTTCATCAGTGGTGGTGGAGACACAGTTTTCGCGATGGCGTC 1320
1357 CGCGCTTGAAGAACCGCGGTGAGATGTCGGGACGTGACGAGTGGGAGTACGTCA 1416
1321 CGCGCTTGAAGAACCGCGGTGAGATGTCGGGACGTGACGAGTGGGAGTACGTCA 1380
1417 AGATCCGAATGCTCAATGACGAGGATGTCATGCTCTGCTTCCCGAGGATTCGTCGCGCT 1476
1381 AGATCCGAATGCTCAATGACGAGGATGTCATGCTCTGCTTCCCGAGGATTCGTCGCGCT 1440
1477 ATGAGATGTTGATGACGCGCATTTCAAGACACGCAACCTCTTGGCAATCTGAAGAACTATC 1536
1441 ATGAGATGTTGATGACGCGCATTTCAAGACACGCAACCTCTTGGCAATCTGAAGAACTATC 1500
1537 TCACACAGGATGTCATCCCGACCTGAAGCGCTTTCAGGATGACGCTCGAAGGCTATC 1596
1501 TCACACAGGATGTCATCCCGACCTGAAGCGCTTTCAGGATGACGCTCGAAGGCTATC 1560
1597 GGGACACGTCATCAGCGGTTTCTCCAAACAGGCGATGTCTGGACACGCGCTCCGATTG 1656
1561 GGGACACGTCATCAGCGGTTTCTCCAAACAGGCGATGTCTGGACACGCGCTCCGATTG 1620

1657 CTAGCGATGGCTGTTTCCAAAGTTTCAGGTGTTCTTGACAGGAAACCGTCGTCGGCGATCG 1716
1621 CTAGCGATGGCTGTTTCCAAAGTTTCAGGTGTTCTTGACAGGAAACCGTCGTCGGCGATCG 1680
1717 AAGACAAGCGGACCTGTACAGTATAGCGTTTCGGAATTCATCTCAAAATGCTGC 1776
1681 AAGACAAGCGGACCTGTACAGTATAGCGTTTCGGAATTCATCTCAAAATGCTGC 1740
1777 GTGTGCGGACGAGAAAGCGGGAGCTATGAATCGTCCGACCGCACTTATGGCGACGCG 1836
1741 GTGTGCGGACGAGAAAGCGGGAGCTATGAATCGTCCGACCGCACTTATGGCGACGCG 1800
1837 AATGGAAGTTGGCCAAAGCGGACGACTTCGAAAGCTCTCTCAAGCTCCCGCGTTCGATG 1896
1801 AATGGAAGTTGGCCAAAGCGGACGACTTCGAAAGCTCTCTCAAGCTCCCGCGTTCGATG 1860
1897 GTGTGCGGACCTGGATACGTTCCGACCTCCGACCTGATGAATCGTCCGACCGCACTTATGGCGACGCG 1956
1861 GTGTGCGGACCTGGATACGTTCCGACCTCCGACCTGATGAATCGTCCGACCGCACTTATGGCGACGCG 1920
1957 TCCGGAAGGAGGCTAAAGCGGCACTCCCGGCTGAATTCGGCTTTTAGGGTAGCGAC 2016
1921 TCCGGAAGGAGGCTAAAGCGGCACTCCCGGCTGAATTCGGCTTTTAGGGTAGCGAC 1980
2017 TGAACAGAAACCGCGCTCTGGAAGGAGCGCGTTCGAAAGCTCTCTCAAGCTCTGTGCCA 2076
1981 TGAACAGAAACCGCGCTCTGGAAGGAGCGCGTTCGAAAGCTCTCTCAAGCTCTGTGCCA 2040
2077 TCAGGACAAGGATCACGACGACCGACGATCAGGACAAGTCCGCTGGAGGGGAGCCCCAT 2136
2041 TCAGGACAAGGATCACGACGACCGACGATCAGGACAAGTCCGCTGGAGGGGAGCCCCAT 2100
2137 TCGAACCTGACGGCCATGACGCGCAGCGACCGAGATCAGGATTAACAAGAGATCAGTCC 2196
2101 TCGAACCTGACGGCCATGACGCGCAGCGACCGAGATCAGGATTAACAAGAGATCAGTCC 2160
2197 CATGACACATCTCTTCCCGGTTGAGACTGGTCTGTGTTCCGGGTGCTAAAAAGTTTC 2256
2161 CATGACACATCTCTTCCCGGTTGAGACTGGTCTGTGTTCCGGGTGCTAAAAAGTTTC 2220
2257 CGTAGGGCGGAGAAATCAAGCTGTCGGTTCGGCTTAATCCGCTCCAGCGCATTTG 2316
2221 CGTAGGGCGGAGAAATCAAGCTGTCGGTTCGGCTTAATCCGCTCCAGCGCATTTG 2280
2317 ATGCGGGCCACCGCTCTGTGCGGTTGCGCTCTGTCTGTGACATAGTTTCTGGGCC 2376
2281 ATGCGGGCCACCGCTCTGTGCGGTTGCGCTCTGTCTGTGACATAGTTTCTGGGCC 2340
2377 AGCAGTCCGGATGATGTTCCGGATCAGGTGCGGACCGGATTCGTGTGTA 2436
2341 AGCAGTCCGGATGATGTTCCGGATCAGGTGCGGACCGGATTCGTGTGTA 2400
2437 GTTGGCTGCGGTTGATCCGAGAAATACGATAGGATCCGCTGTTCCGCTGGCGCG 2496
2401 GTTGGCTGCGGTTGATCCGAGAAATACGATAGGATCCGCTGTTCCGCTGGCGCG 2460
2497 CGATTGTTGCGCTTTTCGGCCGCTTCCTGCTGCGGACCGCAAAATGCCCCGTGA 2556
2461 CGATTGTTGCGCTTTTCGGCCGCTTCCTGCTGCGGACCGCAAAATGCCCCGTGA 2520
2557 ACGGCTCGCAAAATCATGTTTCCCTGCGGTGAAGCTGCGGCTGGGCGCGGATCGGCA 2616
2521 ACGGCTCGCAAAATCATGTTTCCCTGCGGTGAAGCTGCGGCTGGGCGCGGATCGGCA 2580
2617 CGGGCATACGGAACAGTCCGCTCATGAGTTCTCAAGCGCGCGCTATTATCCGCATAG 2676
2581 CGGGCATACGGAACAGTCCGCTCATGAGTTCTCAAGCGCGCGCTATTATCCGCATAG 2640
2677 GCCTTGCCTATTCGCGGGCATACATCTCGAAATCGTCCGCTGGGCGCGGATCG 2736
2641 GCCTTGCCTATTCGCGGGCATACATCTCGAAATCGTCCGCTGGGCGCGGATCG 2700

2737 AACAGCATGCCGACTCTCTTGGTGTATTCGGGGGAACTGGAAGCAGGTCTTGAAGCG 2796
2701 AACAGCATGCCGACTCTCTTGGTGTATTCGGGGGAACTGGAAGCAGGTCTTGAAGCG 2760
2797 TTGATTTTCGTTCCGTTTCCACCGGCGCTGCGATCTTCGCCAGCTTCGCCACAGGGCAACA 2856
2761 TTGATTTTCGTTCCGTTTCCACCGGCGCTGCGATCTTCGCCAGCTTCGCCACAGGGCAACA 2820
2857 AGGCGGATGGCGTAAAGCTGATCTCTTTCGCCAGGCGCGCAGCAATCTTGGCAGCGCG 2916
2821 AGGCGGATGGCGTAAAGCTGATCTCTTTCGCCAGGCGCGCAGCAATCTTGGCAGCGCG 2880
2917 AAAAAGCGCGCTGTTGGGATCGGAGCGGCAATTCGGGGGAAAGCGCTCACTCCAGCG 2976
2881 AAAAAGCGCGCTGTTGGGATCGGAGCGGCAATTCGGGGGAAAGCGCTCACTCCAGCG 2940
2977 CCGCTTGGGCGTTGAGTAGGCAACCGTTATTCGGGGCATGCCCGAGCGCTGCCGCCATC 3036
2941 CCGCTTGGGCGTTGAGTAGGCAACCGTTATTCGGGGCATGCCCGAGCGCTGCCGCCATC 3000
3037 AGTGTCTCCGAAAGGACCAACCGGAAAGCGGACCGGACACCGCAACATCTTGCCTCAG 3096
3001 AGTGTCTCCGAAAGGACCAACCGGAAAGCGGACCGGACACCGCAACATCTTGCCTCAG 3060
3097 ATAGCCATGTCAATCAACCTAGCACCGCGCTCAGACGGGCAAAATGACAGATCGCAGGCTA 3156
3061 ATAGCCATGTCAATCAACCTAGCACCGCGCTCAGACGGGCAAAATGACAGATCGCAGGCTA 3120
3157 GGTGTAGGTGCTGATGCGGCAACCGCGGCTTTCGGGTGCTGAGAGCTAGGAGTTAC 3216
3121 GGTGTAGGTGCTGATGCGGCAACCGCGGCTTTCGGGTGCTGAGAGCTAGGAGTTAC 3180
3217 GAACCTATTCGCTGCTCATGCTTTTGGGCGCAGGCTTCTCTGTTCTGTTTTCATCAGCGAT 3276
3181 GAACCTATTCGCTGCTCATGCTTTTGGGCGCAGGCTTCTCTGTTCTGTTTTCATCAGCGAT 3240
3277 ATTTTATGCCCCCTTGTATCCAGACTGCTACTTTCGATCCCTTTCCGCTCTGATGAGAA 3336
3241 ATTTTATGCCCCCTTGTATCCAGACTGCTACTTTCGATCCCTTTCCGCTCTGATGAGAA 3300
3337 CTGATGATCTTTTGTATCAAGCGCTCCCAATGCTGGCTGCAAGAAAGTCTCAACTGGTTG 3396
3301 CTGATGATCTTTTGTATCAAGCGCTTCGCAATGCTGGCTGCAAGAAAGTCTCAACTGGTTG 3360
3397 CGGGAAGCGGATCATAAATGGGTTTCGATTCGCGGGCGCTGCTGTTTCATCTGGGGCG 3456
3361 CGGGAAGCGGATCATAAATGGGTTTCGATTCGCGGGCGCTGCTGTTTCATCTGGGGCG 3420
3457 GTTCTGCTCATCTGCTGCTGTTCTGGGCTGTTGGGATGCTGCGGCTGGGCTGATGTTGTT 3516
3421 GTTCTGCTCATCTGCTGCTGTTCTGGGCTGTTGGGATGCTGCGGCTGGGCTGATGTTGTT 3480
3517 GCGCAGGATATTCGCTTCTCCGCTCCGCTCAGGCGCGCTTCGCTGCGCTGATCGAAGCT 3576
3481 GCGCAGGATATTCGCTTCTCCGCTCCGCTCAGGCGCGCTTCGCTGCGCTGATCGAAGCT 3540
3577 CAACATCCGATTTGGCTGGGCTTTCCGCGGCAAAAGCGGCAAGCTAACCTTTCTGCTGG 3636
3541 CAACATCCGATTTGGCTGGGCTTTCCGCGGCAAAAGCGGCAAGCTAACCTTTCTGCTGG 3600
3637 ACCTGTTTCTGAAGATGTGTCAGTGTGCAACCGCAGGCTGGAAGCAGTGGCGGCTCT 3696
3601 ACCTGTTTCTGAAGATGTGTCAGTGTGCAACCGCAGGCTGGAAGCAGTGGCGGCTCT 3660
3697 GGTGCTCGCGGCTCGAGAGAACCCAGACAGCAAGCTCTGCTGCGGACTGCG 3756
3661 GGTGCTCGCGGCTCGAGAGAACCCAGACAGCAAGCTCTGCTGCGGACTGCG 3720
3757 GCCATCCGCTCCAGTATAGCCAGAACTCGGGTTCCAGTGGCAGCGCTCCGCTGCTCT 3816
3721 GCCATCCGCTCCAGTATAGCCAGAACTCGGGTTCCAGTGGCAGCGGCTCCGCTGCTCT 3780
3817 GACAGACAGGCTGCGCTTTGACGAGATCACTCATTCGCGTGTGTTTCTCAAGCGCTTCA 3876

||||| 3781 GACAGACAGCGTGGTTCGACGAGATCACTCATTCGGTGTGTTCTCAAGGCGCTTCA 3840
||||| 3877 AAGCCATTGTCGGTTCGCAACATCAGGTCGGGATCACTCAGCAGCTCCCGCGCAG 3936
||||| 3841 AAGCCATTGTCGGTTCGCAACATCAGGTCGGGATCACTCAGCAGCTCCCGCGCAG 3900
||||| 3937 AAGATATAAGCAGGATCGCGCGAGTTCGCGATCGCGATCAGGACAGTACGATACGAAAC 3996
||||| 3901 AAGATATAAGCAGGATCGCGCGAGTTCGCGATCGCGATCAGGACAGTACGATACGAAAC 3960
||||| 3997 CGGTTGCGTCCAATCCGTTTACCGGAGAGCCAGAAAAAAGCTCCGGAATGTCCGCAATTA 4056
||||| 3961 CGGTTGCGTCCAATCCGTTTACCGGAGAGCCAGAAAAAAGCTCCGGAATGTCCGCAATTA 4020
||||| 4057 TCCAGCCGACAGTTCGTCGAGTTTGGTGCATCAGCTCCGCGCGGCGCTG 4109
||||| 4021 TCCAGCCGACAGTTCGTCGAGTTTGGTGCATCAGCTCCGCGCGGCGCTG 4073

RESULT 2
AB063188
LOCUS AB063188 1458 bp DNA linear BCT 19-JAN-2002
DEFINITION Gluconobacter oxydans sr gene for L-sorbose reductase, complete cds.
ACCESSION AB063188
VERSION AB063188.1 GI:18250946
KEYWORDS
SOURCE Gluconobacter oxydans (strain:IFO3291) DNA, clone:8.0 kb EcoRV fragment.
ORGANISM
Gluconobacter oxydans
Bacteria; Proteobacteria; alpha subdivision; Acetobacteraceae;
Gluconobacter.
1
Shinoh,M., Tazoe,M. and Hoshino,T.
NADPH-dependent L-sorbose reductase is responsible for L-sorbose assimilation in luconobacter suboxydans IFO 3291
J. Bacteriol. 184 (3), 861-863 (2002)
21650687
REFERENCE
2 (bases 1 to 1458)
Shinoh,M.
Direct Submission
Submitted (13-JUN-2001) Masako Shinjoh, Nippon Roche Research Center, Applied Microbiology; Kajiwara 200, Kamakura, Kanagawa 247-8530, Japan (E-mail:masako.shinjo@roche.com, Tel:81-467-47-2226, Fax:81-467-45-6812)
Location/Qualifiers
1. .1458
/organism="Gluconobacter oxydans"
/strain="IFO3291"
/db_xref="taxon:442"
/clone="8.0 kb EcoRV fragment"
/note="originally submitted as 'Gluconobacter suboxydans' (synonym: Gluconobacter oxydans subsp. suboxydans)"
1. .1458
/gene="sr"
1. .1458
/gene="sr"
/function="Reduction of L-sorbose to produce D-sorbitol"
/standard_name="NADPH-dependent L-sorbose reductase"
/note="synonym of NADP-D-sorbitol dehydrogenase (NADP-SLDH), homolog of NADP-SLDH with the accession number of AB028937"
/codon_start=1
/evidence=experimental
/transl_table=11
/product="L-sorbose reductase"
/protein_id="BAB83933.1"
/db_xref="GI:18250947"
/translation="MITHETLKSIPAGVQAPPYDINGIKPGIVHFGVGNFFRAHEAFY
VEQILKDDPNWIGTGTGSDRSKKAEFEKQDCLFSLTETAPSKGSTVRVMGALR
DYLLAPADPEAVLKHLPDGPRIYVSMITEGGINNETTGFEDLENKAVQDDLTPTET
PSTIFGVVEGLRRRRONGGKRAFTIMSCDNLNRHNGNARKAFLGAKARDPELAKWIE

ENATFPNGMVDRIPTVSADIKKLNEASGLHDDLPVLAEDFHOWLEDSFADGRPAL
EKAGVQVGVDTYEHVKIRMLNAGHIMLCFPAVLAFENVDHSLADQDRLRLLENFL
NKQIVPTLKKAPPGWTLGYSVTSRESNPAMADQTLRI SDGSKTQVETVRKA
PEKRDLSRLTAFGMAYLEMRKGDGTGYEPEPTFGDNHKTAKADDFESALKLP
AFDNRDLETGSLNNKVELRKIRREKVAALPA"
BASE COUNT 329 a 386 c 434 g 309 t
ORIGIN
Query Match 23.2%; Score 955.6; DB 1; Length 1458;
Best Local Similarity 78.5%; Pred. No. 6e-183;
Matches 1144; Conservative 0; Mismatches 314; Indels 0; Gaps 0;
Oy 537 ATGATTACGGCGGAAACCCCTTAAGTCTCTCTCCCAATGTCCAGGCTCCCGCCCTATGAC 596
Db 1 ATGATCAGCGACGAAACCCCTCAAGTCTCTCCCGCGGTGTGACAGGCTCCCGCCCTATGAC 60
Oy 597 ATCAGCGGGATCAAGCCTGGGATCGTGCATTTCCGTTAGGTAACTTTTTCGAGGCCAT 656
Db 61 ATCAATGGGATCAAAACCGGGATCGTGCATTTTGGCGTGGGAAACTTCTTCCGCGGCCAT 120
Oy 657 GAGCGTTTCTAGCTGACGACAGATTCTTGAACACACTCCGGACTGGCGATTGTTGGTGT 716
Db 121 GAGCTTTTCTAGTTGAACAGATCTCAAGACCATCCGAACCTGGGAATCATCGCGTT 180
Oy 717 GGCCTGACGGCAGTACCGCTTCAAGAAAAAAGCCGAGGAATTCAGAGCCCGAGACTGC 776
Db 181 GGTCTGACGGGTACGACAGGTCAAGAAAGAGCCGAGGAATTCAGAAAGCAGGACTGC 240
Oy 777 CTGTATTCCCTGACCGAGACGGCTCGTCCGGCAAGAGCAGGCTGGCGGTCTATGGCGCG 836
Db 241 CTCTTTTCCCTGACCGAAACGGCTCCGTCGGCAAGAGCAGCGTTCTGTGTATGGCGCG 300
Oy 837 CTGCTGACTATCTGCTTGCCTCCGCGCATCCGGAAGCCGCTGCTGAAGCATCTTGTGT 896
Db 301 CTGAGGATTAACCTTTTGGCTCTCGCGATCCGGAAGCCGCTGCTGAAGCATCTCGTGC 360
Oy 897 CCGGCCATCCGCATCTGTTTCCATCAGCATCAGGAAGCGCGCTACACATCAACAGAGACG 956
Db 361 CCGGAAATCCGTATCTGTTTCCATGACAATCAGGAAGCGCGTTTACACATTAACAGAGACG 420
Oy 957 ACCGCTGCTGATCTGGAGAATCGGCAGTAAGGCCGACCTCAAGAACCCGGAAG 1016
Db 421 ACAGTGAGTTGATCTTGAGAACAAAGCGGTTTCAAGAGGATCTGAAGACACCCGGAACG 480
Oy 1017 CCGTCTACCGTTTTCGGTTAGTGTGAGCGCCCTGCTGCTGCTGCGATGCGCGGTGT 1076
Db 481 CCGTCCACATCTTTGGATATGTTGTGAAGACTGCGCCCGCGCTGACGACGAGTGGC 540
Oy 1077 AAGCATTTAGGTCATCTCTGTGATAACCTGCGTCATACCGCAATGTGCGCCGCAAG 1136
Db 541 AAGGCTTTCAGCATCATCTCTGCGATAATCTGCGGCATACCGTAAATGTGCGCCGCAAG 600
Oy 1137 GCCTTCTCGCTATCGAAGCGCGGATCGGAGTTGGCGAAGTGGATTGAGAGAAAC 1196
Db 601 GCATTTCTGGGATACGCGAAGCGCGGATCCGGAACCTGGCAAGTGGATTGAAGAGAAC 660
Oy 1197 GCGACCTTCCCGAAGCAATGTTGATCGCATCACCCGACCGCTTTCGGCGGAATCGCC 1256
Db 661 GCGAGTTTCCCAATGCGATGTTGATCGCATCAGCGCACCGTTTCTGCTGACATTCGG 720
Oy 1257 AAGAGGTCACACGCGGCGAGTGGCTGTGATGACGACCTGCGCGTGGTGGCGGAGATTTC 1316
Db 721 AAGAAGCTCAACGAAGCAGTGGCTGCACGACGACCTGCGCGTCTGTTGCAAGAGACTTT 780
Oy 1317 CATCAGTGGTGTGGAGACACAGTTTCCGATCGCGCTCCGCGCTTGAAGAACCGCG 1376
Db 781 CATCAGTGGTGTGGAGACACAGCTTTGCTGATGGCGCGCTGCGCTGGAAGAACCGCG 840
Oy 1377 GTGCAGATGTTGCGGGAGCTGACGACTGGGAGTACGTCAAGAGTCCGAATGCTCAATGCA 1436
Db 841 GTGCAGTTCGTTGGGATGTGACGACTACGAGCATGTAAAAATCCGCATCTCAATGCT 900
Oy 1437 GGGCATGTCTGCTCTGCTTCCCGAGGCAATTCGTGCTGCGCTATGAGATGTGGATGACGCC 1496

Db	1261	GGACCTACGAGCATTCGAGCCGACTTTTGTGTGACAAACCATAGACTCTGGCCCAAGGCT	1320
QY	1857	GAGCACTTCAAAGCTCTCTGAAGCTCCCGGCTTCGATGGGTGGCGGATCTGGATACG	1916
Db	1321	GATGATTTTGGAGCGCGCTCAAGCTGCAGCGTTCGATGCTGCGCGATCTGGAGACG	1380
QY	1917	TCCGAACGTGATCAAAAGTCACTGCTGCGGAGATCATCCGCGAAAGGCGGTAA	1976
Db	1381	TCCGGGCTGAACAACAGTTCTGAGCTTCGAAGATTATCCGCGAAGGCGGTCAAG	1440
QY	1977	CGCGCATCCCGGCTCTGA	1994
Db	1441	GCTGCCCTCCGCGCTGA	1458
RESULT 5			
LOCUS	AE004660	10977 bp	DNA linear BCT 30-AUG-2000
DEFINITION	Pseudomonas aeruginosa PA01, section 221 of 529 of the complete genome.		
ACCESSION	AE004660	AE004091	
VERSION	AE004660.1	GI:9948372	
KEYWORDS	Pseudomonas aeruginosa.		
SOURCE	Bacteria; Proteobacteria; gamma subdivision; Pseudomonadaceae; Pseudomonas.		
ORGANISM	1 (bases 1 to 10977)		
REFERENCE	Stover, C.K., Pham, X.Q., Erwin, A.L., Mizoguchi, S.D., Warrenner, P., Hickey, M.J., Brinkman, F.S., Hufnagle, W.O., Kowalik, D.J., Lagrou, M., Garber, R.L., Goltry, L., Tolentino, E., Westbrock-Wadman, S., Yuan, Y., Brody, L.L., Coulter, S.N., Folger, K.R., Kas, A., Larbig, K., Lim, R., Smith, K., Spencer, D., Wong, G.K., Wu, Z., and Paulsen, I.T.		
TITLE	Complete genome sequence of Pseudomonas aeruginosa PA01, an opportunistic pathogen		
JOURNAL	Nature 406 (6799), 959-964 (2000)		
MEDLINE	20437337		
PUBMED	10984043		
REFERENCE	2 (bases 1 to 10977)		
AUTHORS	Stover, C.K., Pham, X.-Q.T., Erwin, A.L., Mizoguchi, S.D., Warrenner, P., Hickey, M.J., Brinkman, F.S.L., Hufnagle, W.O., Kowalik, D.J., Lagrou, M., Garber, R.L., Goltry, L., Tolentino, E., Westbrock-Wadman, S., Yuan, Y., Brody, L.L., Coulter, S.N., Folger, K.R., Kas, A., Larbig, K., Lim, R.M., Smith, K.A., Spencer, D.H., Wong, G.K.-S., Wu, Z., Paulsen, I.T., Reizer, J., Saier, M.H., Hancock, R.E.W., Lory, S. and Olson, M.V.		
TITLE	Direct Submission		
JOURNAL	Submitted (16-MAY-2000) Department of Medicine and Genetics, University of Washington Genome Center, University Of Washington, Box 352145, Seattle, WA 98195, USA		
FEATURES	Location/Qualifiers		
source	1..10977		
	/organism="Pseudomonas aeruginosa"		
	/strain="PA01"		
	/db_xref="taxon:287"		
gene	107..1012		
	/gene="mtlr"		
	/note="PA2337"		
CDS	107..1012		
	/gene="mtlr"		
	/codon_start=1		
	/transl_table=11		
	/product="transcriptional regulator mtlr"		
	/protein_id="AAG05725.1"		
	/db_xref="GI:9948373"		
	/translation="MPRASKVTDPSYELMDDEHSHLIYRQHPSPILVRHFIKEYE LHLIVASSGKVIQDIGNFAPDTLFLGPNLPHNWSIQVEPDEVARMDLVNPTDE VLEDSGVSENLRLAPLARYGIEFRDPALESGRRLLQRIADSGMTRLGFFI LMEQAECDQLLSSTQSLADEHNVRINRAVDYIFQHYAQLDQEOVAHLGMT PTYFRFQKAAAGCFEVEFVNRILRSKCELLARSELPTVEVCESGNSLNSNRRE QOLKGMTPTGYRRLVTRQLTEQNLL"		
gene	1170..2480		
	/gene="PA2338"		

CDS

```
1170..2480
/gene="PA2338"
/codon_start=1
/transl_table=11
/product="probable binding protein component of ABC
maltose/mannitol transporter"
/protein_id="AAG05726.1"
/db_xref="GI:9948374"
/translation="MNDISKACLAACIALPLAQAETFLTITATVNNNDIMRORLSK
VFESHPDIALKWVLEENVRQRLTTDIATGGGQFDLLTGMYEALWKAAGKWLPEM
SGLPADYALDDLPSVRGCLSVKGTLYALPYAEASITYYRKDLFQQAGLRPEQPTW
TOLGEFAARLNRPDQGYGICLRKAGNGENMALIGTANAFGARFWRQPEFSGG
EMKALDYYVSTLRKRYGPGSSNGENLALFNISGKCAIWDASVAGSFYDKSQSK
VDATGFAFAPREVTDKASWLYSWALAIPASSRAKDAKAFATWATKSOAYGKLVADR
EGVANVPCTRASTYSEAYLAAAPFVRTLESKRVDNHPPTLKPVVPGIQLVTYPE
FOAICTOVGKLFSAALTQMSDQALAAQAQSTAREMKRAGYPK"
2556..3488
/gene="PA2339"
2556..3488
/gene="PA2339"
/codon_start=1
/transl_table=11
/product="probable binding protein-dependent
maltose/mannitol transport protein"
/protein_id="AAG05727.1"
/db_xref="GI:9948375"
/translation="MNTSALDTPLPARPARRRQRRRPGMLLVSPSVAILLMMIV
PLGTLVPSLIYRNLVPCENAFVGLNFTFVTVDAGFLPGALNTLLLVGSLVISA
FGVLLAALKEAGFCFGARVRLVLLISPFITMPSALLWKNLFIHPVSGIILAAVMRLF
GAEPVDMUHHYPLLSIILVSNQWMLPFAILMTAMQSDQEQEKAARDGAGPLAIF
WHLPLHARPIAVVLTETIFLLSFAEITFTTSGGPGYESTNLAYLIYNQALLQFD
VGMASAGLIAVLIANIAAIVLVRMIGNKLTERS"
3499..4332
/gene="PA2340"
3499..4332
/gene="PA2340"
/codon_start=1
/transl_table=11
/product="probable binding protein-dependent
maltose/mannitol transport protein"
/protein_id="AAG05728.1"
/db_xref="GI:9948376"
/translation="MLTLKQSRRLRNALLGLLWGLALLFFPIFMMLLTFSKSEIDA
FATPQFOTPLENYLHINERSDYFAYANSLISSATLLCLMVAIPAAYSAFFE
TRHKRTLLMWLSKMLPPVGLMPYLLAKQFGLLDSRLALIVYTLINLPVVMVW
YTYKDIIPVDILEAARDGACTWQEIIVRVLPIARGGLASTLLLSLILCNWEAFSLN
LTSSNAAPLTALVASYSPEGLFWAKLSAVSTLACAPILVFGWISQKOLVRGLSFGAV
K"
```

gene

```
4372..5484
/gene="PA2341"
4372..5484
/gene="PA2341"
/codon_start=1
/transl_table=11
/product="probable ATP-binding component of ABC
maltose/mannitol transporter"
/protein_id="AAG05729.1"
/db_xref="GI:9948377"
/translation="MADLKIRNLQKQEDQALIKGIDLDVRDREVFVVGPSGCGKST
LRLIAGLEASGSGIADLGDITDTPPAKRDLAMVFQTYALYPHMTVRRNLSEALDL
AGVDKREVAARVDAARILEQALERKPOLSGQORVAIGRAIVRNKPIFLFDEP
LSNLDAALRVQMRLAARLHQELATMIYVTHQVEAMTADKVVVLGVRQVQSGP
LELYHIPANLFAVAGFLGTGPKMFLGRHLSRNGERCEVALECGARVGLPLCAGELATG
SPVTLGIRPEHLNIARPCAFQGLLRVAADVSRIGSDSFCHVRADSGEMLTLVRVGD
FAPDYGAPLELGFDPHECHLFDNSQALDKRLQQA"
```

gene

```
5507..6982
/gene="mtlb"
5507..6982
/gene="mtlb"
/codon_start=1
/transl_table=11
/product="mannitol dehydrogenase"
```

CDS

```

/protein_id="AAG05730.1"
/db_xref="GI:9948378"
/translation="MKLNROHLLPLATAVARPSYDPAQLRQGVHVGSGFFHRAHQAA
YTDALMRGALDWAICGALRSDDRAMHDALAAQDLYTLYELGDOPDTEVRVIGAI
SGMLLAEDGAELLEKLAEPARIVLSITITEGVCYIDGSGEFLAELPLVRHDLNPR
TPRGVGLCEALRRDRDTGVPAFTVMSCDNLPHNGVARKALLAFERLDPLGLRWI
ATHVSPNAMDRTITPMKTSAPHRDQLAORHDVEDAMPVCEPFQWVLEDRFSAGRPA
WEKVGQFTDDVTPEEMKIGLLNSHLATYGLFRGVRFVHETLGDPLLRVYRAF
MDRVPAPLAPVPCIDLERYKDSLVERFANRAIADQLERVCSDGSSKPKFIVPTANR
LIAARGRLERVALVVAWALYLGGVDEHGERYPIDPRAAEQCALVAERHGLGRLLIG
VETVFGPAIPASAFVFAFERLDLSLRHGVSETLRQVGE"
6979..8487
/gene="mtly"
/feature="xylulose kinase"
/feature_id="AAG05731.1"
/protein_id="AAG05731.1"
/db_xref="GI:9948379"
/translation="MTPEGRMHGLFGIDCGTQGSKALLLDAGSGRTLGLGSAORPP
EGRDREOPADMLEAMASAVRNALPEAAVDGREVALAVSAQOCHGLLLLDAGRAL
PAKILLDTSAENRELLERAGGAPAGSLERGLVYLPAGYTLKLLNSRRRFPLEPAR
VAHLLPHDYLNHMLTGRVCEADSGSGTFDVRRTWASDVLDELVEPGRLAALP
ELIEPGACIGNLRPEAAALGAPHTRVAGGGDNMLAAGTGNIRPGLLTASLGTSG
TLSAYABRPVSPHGEATFCASSGWLPLACTMNLTCAGLVODLLHLLDDEFSLA
AQAQVAGELMLPFFDGERVAPLPHASASLHGMNTAANLSRANLCRAVLEGTAFGLRY
GLDLRASGLPGEIRLVGGAANKPLNRRTLADLLGLPLVCPRTEAAALQAALQAAW
SLRGESAGESLEALRCVALDESTRTQPARQAAVEQYRYVLEHLPFR"
8528..9460
/gene="mtly"
/feature="xylulose kinase"
/feature_id="AAG05732.1"
/protein_id="AAG05732.1"
/db_xref="GI:9948380"
/translation="MYLVCGEALDFVFSLEAARSNELGFTAIAGGSPFNVAVLRLR
GVEAALFGLSSDYLGRALRLVBEEDVCGFLVPSDAPPTLMAVGLDASGSAOYQR
EGGADROVRLHLEPLDGRIRGLHVGSTYLVTPVADTLLALVRRSGRLVSLDPN
VRLDPQDIDLRRRVEAFASHAHLIKASEEDLALLYPGRDPGEVRLNPRCLRVF
VTHGAGASVHCTHSWSPADTALPLRDTVAGDFTFOAATLAYLRDLADSPAGLAA
LSRAIDAMLAFAIRAAAVTCSRVPDLPFAHEL"
complement(9503..10738)
/gene="PA2345"
complement(9503..10738)
/feature="xylulose kinase"
/feature_id="AAG05733.1"
/protein_id="AAG05733.1"
/db_xref="GI:9948381"
/translation="MQSANASELTDDIAVIGGGAAGTGVIAASLLKRSPLGISLIEPA
LTHYQPGWTLVGGGAYAGDTARPMAGLVPPGVWEVRLRVERDPEARLLLEGGDS
DEYRLNLCPGLRWERIEGLETLGRNGVTSNYRDLAPYTWELVRLGRLLGKALFT
OPAMFKCAGAPKAMYLSCDHLREGVLQDIEVEFDLAGAALFQVADFPVPLMEYVR
KYSABFLNSNLKVDGAARKAWFEVKADNTNIAEKDFLLHVVVPPOLPPTFAAS
GLGDAAGCEVDPATLQHVHGEIFALGDVCGGTANAKTAAARQIIVVAENLLGLRS
GRLPLRDYGGVGLPVTVERGRVLAEEFYAGKLLTPPLEPTAASRFANWLLKRLVLP
WYWNMLKLGREWLPGR"
1674 a 4067 c 3532 g 1704 t
BASE COUNT
ORIGIN
Query Match 8.5%; Score 351.4; DB 1; Length 10977;
Best Local Similarity 55.5%; Pred. No. 8.4e-61;
Matches 726; Conservative 0; Mismatches 571; Indels 12; Gaps 2;
QY 564 CTTCTCGCAATGTCAGGCTCCCGCCCTATGACATCGACGGGATCAAGCTGGGATCGTG 623
II II II II II II II II II II II II II II II II II II II II II II
DB 5537 CTGGCGACGGCTGTGCGCCGCCCTCTACGACCCCGGCCCACTGGCCGAGGCATCGTC 5596

```

```

QY 624 CATTTCGGTGTAGGTAACTATTTTTCGAGCCCATGAGCGCTTCTACGTGCA-----GCAG 677
II II II II II II II II II II II II II II II II II II II II II II
DB 5597 CACATCGGCTCGCGGCTTCCACCGCGCACGAGCGGCTTACACGACGCGCTGATG 5656
QY 678 ATTCTTGAACACCGCTCGGACTGGCGATTGTTGGTCTTGCGCTGACGGGACGACCGGT 737
II II II II II II II II II II II II II II II II II II II II II II
DB 5657 AACCGCGGCGAAGCCCTCGACTGGGCTATCTGGCGCGCGGCTTCGCGACGAGGACCGGC 5716
QY 738 TCAAAGAAAAAAGCCGAGGAATTCAGGCCACGACTGCCTGTATTCCTCCGACGAGACG 797
II II II II II II II II II II II II II II II II II II II II II II
DB 5717 GCCATGCACGACGCGCTGGCC-----GCCAGGACTACCTCTACACCTCTACGAGCTG 5770
QY 798 GCTCCGTCCGGCAAGACGCGGTGCGGTGATGGGCGCGCTGCGTGACTACTCTGCTTGC 857
II II II II II II II II II II II II II II II II II II II II II II
DB 5771 GGCACCAACCGGACACCGAGGTCCGCTGATCGCGCCATACGCGCATGCTCTCGCC 5830
QY 858 CCGGCCCATCGGAAGCGGTGCTCAAGCATCTTGTTCGCGGCATCCGCGCATGTTTCC 917
II II II II II II II II II II II II II II II II II II II II II II
DB 5831 GAGCAGCGCGCGAGCGCTGCTGGAATAACTTCCGAGCGCGGCATCCGCGCATGCTCTCG 5890
QY 918 ATGACGATCACGAAGCGGCTACAACATCAACGAGACGACCGGTGCTTGCATCTCGAG 977
II II II II II II II II II II II II II II II II II II II II II II
DB 5891 CTGACCATCACGAAGCGGCTACTGATCGACATGCGAGCGCGAGTTCCTCTCGCGAG 5950
QY 978 AATCGCGCAGTAAGCGCGGACCTCAAGAACCCGGAAGCGGTCTACCGTTCCTCGGTAC 1037
II II II II II II II II II II II II II II II II II II II II II II
DB 5951 TTGCGCGTGTGCGCCACGACCTGCGCAACCGCGGACCGCGCGGGGTCTTCGGGCTTT 6010
QY 1038 GTGTCGAGGCGCTCGCTGCTGTTGGATGCCCGGTGTAAGGCATTTACGGTTCATGTC 1097
II II II II II II II II II II II II II II II II II II II II II II
DB 6011 CTCTGCCAGGCTCTGCGCGCGCGCGGCGACACAGCGGTGCGCGGCTTACCGCTGATGTC 6070
QY 1098 TGTGATAACCTGCGTCATAAGCGCAATGTGCGCGCAAGGCTTCTCTCGGCTATGCGAAG 1157
II II II II II II II II II II II II II II II II II II II II II II
DB 6071 TGCACAACTCCCGCATACGCGGAGTGTGCGCGCAAGCGGTGCTGCGGCTTCCCGCAA 6130
QY 1158 CGCGCGCATCGGAGTGTGCGCAAGTGTGATTCAGAAACGCGACCTTCCCGCAACGGAATG 1217
II II II II II II II II II II II II II II II II II II II II II II
DB 6131 CGGCTCGACCGCGGCGCTGGCGCGTGTGATCGCGACGACGCTTCCCGCAACGCGCATG 6190
QY 1218 GTTTCGATCATCACCGCGCGCTTTCGCGGAAATCGCAAGAAAGCTCAACCGCGCGCATG 1277
II II II II II II II II II II II II II II II II II II II II II II
DB 6191 GTCACCGCATCACCGCGCATGACGACCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 6250
QY 1278 GGGCTGGATGACGACCTGCGCGCTGCTGCGCGGAGGATTTTCATCATGAGTGGGTGCTGGAAGAC 1337
II II II II II II II II II II II II II II II II II II II II II II
DB 6251 GACGTGGAAGACGCTTGGCGGCTGCTGCGCAACCTTCTGTCGCAATGAGTGGTGTGGAAGAT 6310
QY 1338 CAGTTTGGGATGCGCGCTCCCGCGCTTCAAAAGCGCGCGTGCAGATGTCGCGGACGCTG 1397
II II II II II II II II II II II II II II II II II II II II II II
DB 6311 CGTTTCAGCGTGGCGCGCGCGCTTGGGAAAGGTGCGGCTGCGAGTTCACCGACGACGCTC 6370
QY 1398 ACGGACTGGGATGCTCAAGATCGCAATGCTCAATGAGGCGCATGTGCTGCTGCTTC 1457
II II II II II II II II II II II II II II II II II II II II II II
DB 6371 ACGCGCTACGAGGAGATGAAGATCGGCTGCTCAACGCGACCGCGCGCGCGCGCGCG 6430
QY 1458 CCAGGCAATTCGTGGCTATGAGAATGTGATGACGCGCATTAAGACAGGCAACTCCTT 1517
II II II II II II II II II II II II II II II II II II II II II II
DB 6431 CTGGGCTTCCTGCGCGCTACCGCTTCTGCTCCACGAGACCGCTCGCGGACGCGCTGCTGCGGC 6490
QY 1518 GCAATCTGAAGAACTATCTCAAGAGGATGTATCCCGACCGCTTCAAGGCGCGCTTCAAGGC 1577
II II II II II II II II II II II II II II II II II II II II II II
DB 6491 CGCTACGCTGGCGCGCTTCATGACCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 6550
QY 1578 ATGACGCTCGAAGGCTATCGGACAGCGCTCATAGCCGCTTTCCTCAACAGCGCATGTCG 1637
II II II II II II II II II II II II II II II II II II II II II II
DB 6551 ATGACGCTCGAAGGCTATCAAGACAGCGCTTGTGAGCGCTTTCGCGCAACCGCGCGCATGCC 6610
QY 1638 GACGACGCTCGGATGCTGAGCGATGCTGTTTCCCAAGGTTTCAAGGTTTCTGAGCGAA 1697
II II II II II II II II II II II II II II II II II II II II II II
DB 6611 GACCAACTGGAACGCGTCTGCTCGGACGCGCTGCTGCAAGTTCCTCAAGTTCATCTGTCGCC 6670

```

QY 1698 ACCGTGCGTGGCGCATCGAAGAACGCGGACCTGTCACTATAGCGTTCGGAATTGCA 1757
 Db 6671 ACCGCCAATCGCTGACCGCGCGCGCGCTGGAACGCGTGGTGGTGGCG 6730

QY 1758 TCCTATCTCGAAATGCTCGTGGTGGCGGACGAGGCGGACGATGATGATCTCCGAG 1817
 Db 6731 CGCTGGGCGCTGTACCTCGGCGGGTGCACGAGCATGCGGAACGCTACCCGATTCGCCGAT 6790

QY 1818 CCGACTTATGCGCGCGCATGGAAGTTGCCAAGCGCGGACGACTTCG 1866
 Db 6791 CCGCGCGCGCGCATGGAAGTTGCCAAGCGCGGACGACTTCG 6839

RESULT 6
 AF007800
 LOCUS
 DEFINITION
 Pseudomonas fluorescens mannitol operon, MtlE (mtlE), MtlF (mtlF),
 MtlG (mtlG), MtlK (mtlK), mannitol dehydrogenase (mtlD), xylulose
 kinase (mtlY) and fructokinase (mtlZ) genes, complete cds.
 AF007800.1 GI:2293413

ACCESSION
 VERSION
 KEYWORDS
 SOURCE
 ORGANISM
 Pseudomonas fluorescens.
 Bacteria; Proteobacteria; gamma subdivision; Pseudomonadaceae;
 Pseudomonas.

REFERENCE
 1 (bases 1 to 10031)
 Brunker, P., Altenbuchner, J., Kulbe, K.D. and Mattes, R.
 Cloning, nucleotide sequence and expression of a mannitol
 dehydrogenase gene from Pseudomonas fluorescens DSM 50106 in
 Escherichia coli
 Biochim. Biophys. Acta 1351 (1-2), 157-167 (1997)

JOURNAL
 MEDLINE
 PUBMED
 97236441
 98121321
 9461423

REFERENCE
 2 (bases 1 to 10031)
 Brunker, P., Altenbuchner, J. and Mattes, R.
 Structure and function of the genes involved in mannitol, arabinol
 and glucitol utilization from Pseudomonas fluorescens DSM50106
 Gene 206 (1), 117-126 (1998)

JOURNAL
 MEDLINE
 PUBMED
 98121321
 9461423

REFERENCE
 3 (bases 1 to 10031)
 Brunker, P., Altenbuchner, J. and Mattes, R.
 Direct Submission
 Submitted (10-JUN-1997) Institute of Industrial Genetics,
 University of Stuttgart, Allmandring 31, Stuttgart 70569, Germany
 On Aug 4, 1997 this sequence version replaced gi:2065484.

COMMENT
 Location/Qualifiers
 1..10031
 /organism="Pseudomonas fluorescens"
 /strain="DSM50106"
 /db_xref="taxon:294"
 814..842
 932..2242
 /gene="mtlE"
 /gene="mtlE"
 /note="similar to E. coli periplasmic maltose binding
 protein MaltE; periplasmic subunit of the periplasmic
 binding protein dependent mannitol transport system"
 /codon_start=1
 /transl_table=11
 /product="MtlE"
 /protein_id="AAC04468.1"
 /db_xref="GI:2293414"

translation="MKFAKALLACTCMTLSAVSLGAQTLTATVNSDMIRMOKLSK
 TFEAHPKILNWWVLENLVRQLTTDIATGGQFDVLTIGMYEALWGAAGWLEPM
 KDLPAVDLDFVSVRDLGSLVAGLAPYAESSTYRTDLFKDAGLSMPHPTW
 SQGEFAALDKSKEQYGLCLRGKAGWENMALITILANGYGARWFEKWPQFNG
 EQDALNFDNMKSGPPGASSNGENLALFNSGKCAIWDVASVAGSFVTDKTSK
 VADHVGTFEAPHEKTDKTSWLSLAIPSSKAKDAKAVFTSWATSKYEGALVAKT
 DGVANPPGTRKSTYSDEYMAKFAFVLTLESKLKADPTKLPKPIGIGLIQVTYPE
 FOAIGTQVGRFFSGALTGQQTVDAAALTAQAQTTERMKRAGYKP"

gene
 CDS
 gene
 CDS
 gene
 CDS

2381..3271
 /gene="mtlF"
 2381..3271
 /gene="mtlF"
 /note="similar to E. coli MaltF protein; inner membrane
 subunit of the periplasmic binding protein dependent
 mannitol transport system"
 /codon_start=1
 /transl_table=11
 /product="MtlF"
 /protein_id="AAC04469.1"
 /db_xref="GI:2293415"

translation="MNTSAKNRLANPGMFLVTPSVALLLLMMIPLVGLMTLVFSIRYN
 LLYPENQFVGLNFTYFIDSGFLPGATNTLLVGSVLISVYVGLISALLEASEF
 LGRGIVRVLISPFIMPTVGALIKWNLIFHPVSGILAAVWKFAGCPVDMLAHPIL
 SVITISVOMWLPFAILLMTAMOSLDQEKAAERLDGAGATAIFMHLTLPHLARPIAV
 VMIETLILSVFAEIEFTTNGGPGYASTNLAYLIYNQALVQFDVGVASAGLLAVVI
 ANIAAILVRMIGRNLTKDP"
 3282..4112
 /gene="mtlG"
 3282..4112
 /gene="mtlG"
 /note="similar to E. coli MaltG protein; inner membrane
 subunit of the periplasmic binding protein dependent
 mannitol transport system"
 /codon_start=1
 /transl_table=11
 /product="MtlG"
 /protein_id="AAC04470.1"
 /db_xref="GI:2293416"

translation="MTLQSSRRSLLGLTLAWAIAILIFFPIFMVLTSPKTEIDAF
 ATPQFIFTPTLENYLHINRSNYFSYAWNSVLISFSATALCLLISVPAAYSMAYET
 KRKSTLLMLSTKMLPPGVLMPIYLLAKSGLLDTRIALIITILINLPVVMVY
 TYPKDIPKIDLEAARLDGATLQOMRVLLLPKAGGLASTVLLSLILCWNEAPWSLNL
 TSSNAAPLTALIASYSSPEGLFWAKLSAVSLACAPILIFGWISQKQLVRGLSFGAVK
 "

4136..5239
 /gene="mtlK"
 4136..5239
 /gene="mtlK"
 /note="similar to E. coli MaltK protein; ATP binding
 protein subunit of the periplasmic binding protein
 dependent mannitol transport system"
 /codon_start=1
 /transl_table=11
 /product="MtlK"
 /protein_id="AAC04471.1"
 /db_xref="GI:2293417"

translation="MANLKIKNLQKGFSGSIKIGIDLEVNDFVVFVPGSGCGKST
 LLRLIAGLEEVSEGTIELDGRDITEVTPAKRDLAMVQFYALYPHMSVVRKMSFALDL
 AGVDKLVESKVNAAARILELGPPLERKPKLGGQQRVAIGRAIVRNPKIFLFDPEP
 LSNLDAALRVQMRLELARLHKLQATMIYVTHDOVEAMTLADKVVVNLNSGRIEQVSP
 LELYHPANLNVAGFLGTPKMGFLKGVTRVDGCGCEVQLDAGTLISLPSGLSLSVG
 SAVTLGIRPEHLEIASPGQTTLTVTADVGERLSDGTFCFHVITSNGEPLTMKIRGDMAS
 QYGTLLHLDPAPACHLFDTDGVAVAVPLRAA"
 5258..6739
 /gene="mtlD"
 5258..6739
 /gene="mtlD"
 /EC_number="1.1.1.67"
 /note="MtlD"
 /codon_start=1
 /transl_table=11
 /product="mannitol dehydrogenase"
 /protein_id="AAC04472.1"
 /db_xref="GI:2293418"

translation="MKLNKQLTALPEVKLPAYTLADTROIAGIHVGGFPHRAQAY
 YTDALMTGEGLDWSICGSLRDRKARDLAGODYLFTLYELGDDTDEVRVIGSI
 SDMLLADSQAALDKLSPERIRVSLTITEGGYCIDDSNGEFAHLPOIHOHLAHPIS
 SPKTVFGICAALQRRAGGIPAFVMSCONLPNGAVTRKALLAFAALHNAELHDMWI
 KAHVSFNAMVDRITPMTSTAHLQLHDEHGIDDAWPVCEPFWQVWLEDKFVNGRPA
 KHVGVGFTDDTVDVTEEMKICLLNGSHLALTLYGLFKGYRFVHETMMDPLFVATMAY
 MLDVTPNLNAPVPGIDLTQYKQTLVDVRSNOAIADQLERVCSDGSSKFPKFTVPTINR


```

PVSCAAEVSRQIAEAVFAAMVQALPKVTAAPAGSSGNFALGNDPARGRDYVMYQI
SGGGYNGNSHGDITNGCTTIGSKSPPEIMEAQFVPLVRYHALRSGSGVAGKRRGG
FGLAYEVEILHGDARAFVMDHGRFGQALGKGDGPNSTVTFRGGVAHVPHLSXE
QDILPKAGDRVRVCTPGGGYGDPRERDADQVQDVALGYTAEEASEKFCGVLSADG
LTVRKATDGGORAG"
10720..11493
/gene="mlr4789"
10720..11493
/gene="mlr4789"
/codon_start=1
/transl_table=11
/product="amino acid ABC transporter, periplasmic amino
acid-binding protein"
/protein_id="BAB51366.1"
/db_xref="GI:14024764"
/translation="MRLFGREVLAAVALTLASGAASQDKLRIGTEGAYPPENYANA
DGTGGDDIDLGKALCAEMKAECEFTVQDFGSPALQAGKFDALINHTIPERAKEY
EFTHYOTPAITVPKDDTLGATGSDLLKGLVQVQTAHQFASQKSGSVTKAY
PTGDARTDMANGRLADMDGSIILTEMLKPKDGCCKLLGLTADPAIHGPGVIAL
QKNGELAGKFNAAIDALRANGKYKEINDKYFSEFDVGN"
11674..12801
/gene="mlr4790"
11674..12801
/gene="mlr4790"
/codon_start=1
/transl_table=11
/product="aminomethyltransferase"
/protein_id="BAB51367.1"
/db_xref="GI:14024765"
/translation="METFLAERRRPFYSSIVGLGATMGRVGGDFISAKYSGVTDHHL
NTRAVGQDLSLTKMDIKGPDAAELVNHVINDAVAMKPGQVRYSTVREDGGIN
DLTVRFGPEHMLVTSVNRKMLPWLHQAQKAYVTDITAAVAPPTIQGPSRSE
LLKAMISADIDKXLRWAFTHSVNGTRVLTSRGTGTGELGFLVPADEAASVMDTL
MRAGDGLKPYGLWMTGLGLEKAYPAHJDMDETPTPHVGLDRWKFKDKGDFIGR
EALLKIRKGDIDERTWTLGILDGNKPAATDARVLADGEDAGITVYSDHGYSLGKVLATA
HLRPLPFAIGTELSIDIGRPTRAVAVAPMPEFDFEGVLRG"
11674..12955
/gene="mlr4791"
11674..12955
/gene="mlr4791"
/codon_start=1
/transl_table=11
/product="ABC transporter, permease protein"
/protein_id="BAB51368.1"
/db_xref="GI:14024766"
/translation="MSPILFELLRSIWETVLTMAASGLISLVLPLCLALISTERG
7.1%; Score 292.4; DB 1; Length 332635;
Best Local Similarity 55.1%; Pred. No. 5e-49;
Matches 616; Conservative 0; Mismatches 496; Indels 6; Gaps 2;

QY 551 AACCTTAAGTCTCTCTCGCAATGTCCAGGCTCCCGCTTATGACATCGACGGATCAA 610
DB 107979 AAACCTTCACGCTCCCGCAATGTTCGCGTCCGGAATACGACCGCTCGAAGCTCAA 107920

QY 611 GCCTGGGATCGTGAATTCGGTGTAGGTAACTTTTTCGAGCCCATGAGCGGTTCACGT 670
DB 107919 GCGCGCATCGTGAATTCGGCTCGGCAATTTTCATCGCTCGCACAGCGGTCTACCT 107860

QY 671 CGACGAGATCTTGACACGCTCCGGACTGGCGATTTGTTGGTGTGGCTGACGGGCGAG 730
DB 107859 CGACGATCTGTTCAATTCAGGTGTGGCCATGACTGGCGCGTGTGTCGGCGAGCGGTGT 107800

QY 731 TCACCGTTTCAAGAAAAAGCCGAGGAATTCAGGCGCCAGGACTCCCTTATTCCTTCAC 790
DB 107799 CGAGGGCAGAGATCGTTCGCGCAAGCTCGAGAGCAGGACTGGCTGACCCAGGTGT 107740

QY 791 CGAGACGCTCGCTCCGCGCAGACAGCGGTGCGCGTATGCGCGCTGCTGCTGACTATCT 850
DB 107739 CGAGCAGAT---GAAGGCCATATGAGCGCCGCTGTCTACCGCGCGATGATCGATTTCT 107683

QY 851 GCTTGCCCGCGCATCGGAGCGGTGCTGAGCATCTTTGTTGATCGGCGCATCCGCAT 910
DB 107683 GCTTGCCCGCGCATCGGAGCGGTGCTGAGCATCTTTGTTGATCGGCGCATCCGCAT 910
```

```

Db 107682 GAT---CCCTGGCGATCGCGCGGCATCATCGACCGTGTGGCCGACCGCGCATCAAGAT 107626
QY 911 CGTTTCCATGACGATCACGAAGCGCGCTCAACATCAACGAGACGACCGGTGCTGCTCGA 970
DB 107625 CGTTTCCGTCGACATCACGAAGCGCGCTATTTTCATCGACCGCGGTGCGCAAGTTCAA 107566
QY 971 TCTGAGAAATGCGGCAGCTAAAGCGCGACCTCAAGAAACCCGGAAGACCGCTTACCGTWT 1030
DB 107565 CCAGGCCATCCGGATATTCGTGCGCGACGCCAACCCGGCGGCCACCCCAAGACAGTGT 107506
QY 1031 CGTTTACGTGTCAGGCCCTCGCTGCTGCTGGATGCGCGGTGTRAGGCATTTACGGT 1090
DB 107505 TGGCATCATCTGCGAGGACTGCTGCGCGCGCGCAAGGACCGCTGCGGTTCACCGT 107446
QY 1091 CATGCTCTGTATAACCTCGCTCATAAAGCAATGTGCGCGCAAGCGCTTCTCCGCGTA 1150
DB 107445 GATGCTCTGCGACAACATTTCCACACAGCGGCATCTCACCTCGGATGGCTCATTTGCT 107386
QY 1151 TGCGAAGCGCGCGCTCGGAGTTGGCGAAGTGTGAGGAAACCGGACGCTTCCCGAA 1210
DB 107385 GCGCGCGCTGATCATGAGGATTTGGCGGTGGTGGTGGAGCAACGCTGCGCTTTCAAA 107326
QY 1211 CGGAATGTTGATCGCATCACCGCGACCGTTCGCGGGAATCCGCCAAGAGCTCAACGC 1270
DB 107325 CGCATGTTGGACCGCATTTAGCGCGGCCACACAGACCGCGAGCGGCATCTACTGGCAA 107266
QY 1271 GCGCAGTGGGTGATGACGACCTGCGCTGCTGGCGGAGGATTTCCATCATAGTGGGTGCT 1330
DB 107265 GGAATTCGGCTCGAGGATGCTGCGCGGTGTTCTGCGAGCGGTTCGCGCAATGGGTGCT 107206
QY 1331 GGAAGACAGTTTCGGATGCGCGCTGCGCGCTTGAAGAACCGCGGTGCGAGATGGTGG 1390
DB 107205 GGAAGACCATTTACCGCATGCGCGCTGCGCGCTGGAAGAGTGTGATAGTTTCTGTCAG 107146
QY 1391 GGAGCTGACGACCTGGGAGTACGTCAAGATCCGAATGCTCAATGAGGCGATGTCATGCT 1450
DB 107145 GGATGTCGCGCGCTTACGAGTTGATGAAGATCCGATCTTCAATGGCGGCCATGCCAGT 107086
QY 1451 CTGCTTCCAGGCAATTCGTGCGCTATGAGAAATGTGGATGACGCGCATTTGAAGACAGCGA 1510
DB 107085 CGCCTATCCGCGCGGTGATGAGATCCATTTCTGTCATGAGGCGGTGAGGAGCGCT 107026
QY 1511 ACTCCTTGGCAATCTGAAGAACTATCTCAACAGAGTGTATCCCGACCTTGAAGGCGCC 1570
DB 107025 GGTGCGCGGTCTTCTGACAAAGCTCGAGCAGCAGAGATCATCCGCGGTGCGCGCGT 106966
QY 1571 TTCAGGATGACGCTCGAAGGCTATCGGACAGCGTATCATGCGGTTCCTCCAAACAGGC 1630
DB 106965 GCGGACAGCGTGTGAGGATTTATACCAGCTCATCGAAGGCGCTTTTCCAAACCCAA 106906
QY 1631 GATGTCGACACGACGCTCCGATTTGCTAGCGATGGCT 1668
DB 106905 GATGCGGACACCATCCGCGCGCTCTGCTCGACGGCT 106868
```

RESULT 8

SME591790/c

LOCUS

DEFINITION

ACCESSION

VERSION

KEYWORDS

SOURCE

ORGANISM

REFERENCE

AUTHORS

TITLE

SME591790 323450 bp DNA linear BCT 05-JUL-2002
Sinorhizobium meliloti 1021 complete chromosome; segment 9/12.
AL591790 AL591688
AL591790.1 GI:15075230
Sinorhizobium meliloti.
Sinorhizobium meliloti
Bacteria; Proteobacteria; alpha subdivision; Rhizobiaceae group;
Rhizobiaceae; Sinorhizobium.
1 (bases 1 to 323450)
Capela,D., Barloy-Hubert,F., Gouzy,J., Bothe,G., Ampe,F., Batut,J.,
Boistard,P., Becker,A., Boutry,M., Cadieu,E., Dreano,S., Gloux,S.,
Godrie,T., Goffeau,A., Kahn,D., Kiss,E., Lelaure,V., Masuy,D.,
Pohl,T., Portetelie,D., Puehler,A., Puehler,B., Ramspeiger,U.,
Renard,C., Thebaud,P., Vandenbol,M., Weidner,S. and Galibert,F.
Analysis of the chromosome sequence of the legume symbiont

JOURNAL MEDLINE PUBMED REFERENCE AUTHORS TITLE JOURNAL	Sinorhizobium meliloti strain 1021 Proc. Natl. Acad. Sci. U.S.A. 98 (17), 9877-9882 (2001) 21396507 11481430 2 (bases 1 to 323450) Gouzy,J. Direct Submission Submitted (26-JUL-2001) Gouzy J., Submitted on behalf of the MELILO EU Consortium MELILO EU Consortium: Laboratoire de Biologie Moléculaire des Relations Plantes-Microorganismes, UMR215-CNRS-INRA, BP27, F-31326 Castanet, France, Laboratoire de Genetique et Developpement UMR6061-CNRS, Faculte de Medecine, 2 avenue du Pr. Leon Bernard, F-35043 Rennes, France, GARC GmbH, Fritz-Arnold-str. 23, D-78467 Konstanz, Germany, Universitaet Bielefeld, Biologie IV (Genetik) Universitaetstr 25, D-33615 Bielefeld, Germany, Unite de Biochimie physiologique, Universite Catholique de Louvain, Place Croix du Sud 2, Bte 20, B-1348 Louvain-la-Neuve, Belgium, Unite de Microbiologie, Faculte des Sciences Agronomiques de Gembloux, Avenue Marechal Juin 6, B-5030 Gembloux, Belgium. E-mail:Jerome.Gouzy@toulouse.inra.fr http://sequence.toulouse.inra.fr/meliloti.html.	gene CDS	KAHLSGAEGTTCALLDRRACTIQIOLAYARGLARALAAGAIEFTDPLLAASRQGLD WNLKTPRGTVTHARVILATNAYGSLVGVPMKEYROELTILPYFOFATNPLPDNVAAR ILPEROGAWDTGLVMTSEFMDRONRLIFGSGIRLDAIAAGTHRAFAARSVRKLPFYIG DFEEHWDGRIGMTTNNLPAHVLAPNVYSISYNGRGTAPGTVFGRALARHVTGDT SAIPLAETVPVDPWPRTLSAFYHAGAQAHHFIDKRF" complement(3024..3740) /gene="SMc01579" complement(3024..3740) /gene="SMc01579" /function="miscellaneous; hypothetical/global homology" /note="Product confidence : hypothetical Gene name confidence : hypothetical predicted by Codon_usage predicted by Homology predicted by Framed" /codon_start=1 /evidence-not_experimental /transl_table=11 /product="CONSERVED HYPOTHETICAL PROTEIN" /protein_id="CAC46789.1" /db_xref="GI:15075233" /db_xref="SPTREMBL:Q92NJ2" /translation="MAVITANAKGGAGKTTAAILSTELARQGNRVVVLDADPQWVI TSWSEVSGRVANLEIVSHVTPASLPCHIRELKEADFIIDLAKDAIVALALGUSD HVLIPVQGCAMDARGAVQIILELRHIGEKARINSHSVLTNRVSLVTTTRALQTIKAL LASRGVSLDTPIVERVAYREIFECCGTLQMDPNRVSNLDKARENAYALAAAEVQNLL PVTARRALMSRLRSALPRAA" complement(3949..5094) /gene="SMc01582" complement(3949..5094) /gene="SMc01582" /EC_number="1.1.1.-" /function="small molecule metabolism" /note="Product confidence : putative Gene name confidence : hypothetical predicted by Codon_usage predicted by Homology predicted by Framed" /codon_start=1 /evidence-not_experimental /transl_table=11 /product="PUTATIVE ALCOHOL DEHYDROGENASE PROTEIN" /protein_id="CAC46790.1" /db_xref="GI:15075234" /db_xref="SPTREMBL:Q92NJ1" /translation="MTITANKSVYKFCAGRIKELADHCKALGKPKLLVTDGRLA PMATQALDILEAGGLGRAIFADVPDNPNDRLNLAGVKAFERDGGHGVVAFGGGSL DLKCVAFMAGQTRPVWDFEDIGDWMTRASVEGIPIVAVPTTAGTGEVGRASVITN SASHVKKVIIFHPKFLPGVTICDPELVGMPKVITAGTMDAFACHLEAYSSPFYHPMS AGIALGCMRLVKYLLPRAYKDGLDLARANMMSAAAMGAVAFQKGLGAIHLSHPVGA IYNTHGMTNAVVMPPVLRNFRSAIEEKIGRAAYILGAGGFGDYVLRLREELGV PDKLSALGVGCTDRIDEMAENAIVDPTAGGNPVELTLDAAEKLEAECI" complement(5091..6476) /gene="SMc01588" complement(5091..6476) /gene="SMc01588" /EC_number="1.2.1.-" /function="small molecule metabolism" /note="Product confidence : putative Gene name confidence : hypothetical predicted by Codon_usage predicted by Homology predicted by Framed" /codon_start=1 /evidence-not_experimental /transl_table=11 /product="PUTATIVE ALDEHYDE DEHYDROGENASE PROTEIN" /protein_id="CAC46791.1" /db_xref="GI:15075235" /db_xref="SPTREMBL:Q92NJ0" /translation="MTMIRCVSPVDGEVYAEIPAIPLEMARQAVAHARLAQKAWARRP LDERVLKVLVAGVARLNEVDEVPQVPELAWQMRPVRYGGEKFGFNEKRSYTAADAL KPLVVEESDRFERRIAREPHGVVFIAPNMPYMTAINTVAPALMAGNTVILKHSQT
FEATURES source	1..323450 /organism="Sinorhizobium meliloti" /strain="1021" /db_xref="taxon:382" 155..1519 /gene="SMc01575" 155..1519 /gene="SMc01575" /function="miscellaneous; hypothetical/partial homology" /note="Product confidence : hypothetical Gene name confidence : hypothetical predicted by Codon_usage predicted by Homology predicted by Framed" /codon_start=1 /evidence-not_experimental /transl_table=11 /product="HYPOTHETICAL PROTEIN" /protein_id="CAC46787.1" /db_xref="GI:15075231" /db_xref="SPTREMBL:Q92K34" /translation="MTPATRTGLSLFAALAMELAPANAQDAYGGYGGGVMLVT PEGDILDYIPGEAEVHAMDRGRGTVLVDPMGNIVATVPNDGVRGQRRYGRDGY PEPRDRGYGSEPGCTGAIPYERDIAPEVEREDLPNSLPISLDREAAAYDPQYDDP LAQPMPTVTCGRARAEIALQVFLDREGFSFGVIDGKGNVTKAIEAWQATGET LDNNPEDIILRFRNGGLITITTTIADAAAGFFVASIPEDTAHRAQLPHLSFTSVT EMLGEKPHMDEAYLRELNPQGVDSIPTTILKVVNPGNKKGVARIADVAKRQVLA DEAGKLIATYPTSTIGSSDTPSPSGTVHERIAFDPPGYTNPKINFOQGANDRILQLOP GPNPGVTVMIALSKPTYIGHGTPEPSKIGKTQSHGCVRLTNMDATELGMKMGSTGVTV EFVD" 1707..2996 /gene="SMc01576" 1707..2996 /gene="SMc01576" /function="small molecule metabolism" /note="Product confidence : putative Gene name confidence : hypothetical predicted by Codon_usage predicted by Homology predicted by Framed" /codon_start=1 /evidence-not_experimental /transl_table=11 /product="PUTATIVE OXIDOREDUCTASE PROTEIN" /protein_id="CAC46788.1" /db_xref="GI:15075232" /db_xref="SPTREMBL:Q92NJ3" /translation="MATERALPNLWHATAPAPRTAPLAGDLTVEVAIVGGFTGLSA ALHAEETGRTAVIEARMIGFGSGRNVGLNVMVQPPDLDLIATIGAAAGNRLDLEL GDGPFYVDLVARHGIECEAVRNGTLHMSVAGBLKEIREAREQWKNRGAPVEVLSAE	gene CDS	


```

/organism="Rhodobacter capsulatus"
/strain="SB1003"
/db_xref="taxon:1061"
/chromosome="1"
complement(<1..456)
/codon_start=1
/transl_table=11
/product="CobD protein"
/protein_id="AAC16109.1"
/db_xref="GI:3128257"
/translation="MSDLRDTATGOAGCPETAPSVRLAAEAKAKATAQGTGLFLLAL
MLFGLMDVAAKLADYPPAMVYVWAFVRLVLLFGLKATLVRVSOQPLQRLG
CFQMGTVLFFTAIRYIGLADAAALGALSPVLITLGAALFUGERIGAGRI"
complement(453..2471)
/EC_number="2.2.1.1"
/codon_start=1
/transl_table=11
/product="transketolase"
/protein_id="AAC16110.1"
/db_xref="GI:3128258"
/translation="MDLAALRAKTPDHKWLATAIRVLAIDVAQAAASGHPGMPGMAD
VATVFLRHLPKFDKAPNWDREVLVSAGHGMILLYALLHLTGVEQATLDEVKFRO
WGRMAGHEXGLEGVETTCPLGGISTAVGMAIAEKMAARECKLVBHKIWIYA
GDCLMGIESQEAIGLAKQELDNILVLDNNNTIDGKVTVDSDVDKARFAASGWD
VUSCGDAEDIDRALTAAKKARPVLDVCKTLIGFSPNKADSYAVHGPGLDGEIK
LTREAYGHEGFVIPAETKAEMEAIGAKGAERAEWEARLAALPAGKRAEFERQMAR
GVAPKLGAIRAFKAQSEAPKVAATKASEMLVAANVPVPTIGGSADLTGSNLTK
TSDIEDFWGNHKGMYRGIREHMAAAMGMWLGVRVPGYGFCTDYARGAMR
LSSLMGVTVVYMTSDISGLGEDPTHOPVEHLAICRATPNTWTPADVPDYTAEAW
LALSSEPPSVLALSRONLPTLRTHKAKNLTKAGYVIAEAEGRQAILMWTGSEVE
IALKARALLQEAIGTRVYVMPCEMFELFAQDEAYRKRILPAGGVRAVEAAIROPWDR
WLLGRGMEKAGFVGMGFGASPAERLYAEFFGTPEAIAAKVKSLI"
complement(4192..4644)
/codon_start=1
/transl_table=11
/product="hypothetical protein"
/protein_id="AAC16111.1"
/db_xref="GI:3128259"
/translation="MITOKMKYALKALLEADEARRDSPQALITBEIATRSQTPKRRFL
EHLIDLRKAGVASIRSGSGYSLKAPROVDSIELVRRIDGPIAPLPCLSRSAYQP
CDCHDEAARCKRVFAEVFWSYLVLSLTDLEMLAAGNPAKLLD"
4869..5441
/codon_start=1
/transl_table=11
/product="hypothetical protein"
/protein_id="AAC16112.1"
/db_xref="GI:3128260"
/translation="WTLAAFSIALLVLLAPGPTNTLMVAGASHCLARVLVPAEL
AGYLLTVVPLAAGLWARAPGAALLKLAALWVWLAVRLWCSAAGREGFAIGA
GRIFVTITLNPALIFGLVLLPAPSPAETAAKLLVFCALVAVALLWGGFGALSCHARA
GEARAGGLRRITASGWLALVSLSLLGLRG"
5513..6715
/codon_start=1
/transl_table=11
/product="potential leu/ile/val-binding protein precursor"
/protein_id="AAC16113.1"
/db_xref="GI:3128261"
/translation="MPKPLVLAARLLTLCALVSGLGSTAPVAAQQAQPGGDLAFAQVAP
LAGPAAALGTGNTGIRAAFEANRAGVGKRLKLEAFDDGYBETSAERTAVIAG
NAHLALIGAVCTPTQATQPLAERAGLPFIGFTCAAFLRQPLRNVNLRPSGAEAA
EAWOHLVETRASRSRSLYQDDSGFRAGLALEAKRLQLTARGAYVNTLVAK
TALLGRAGVEAVVLVGAYPKVAEFARVAHRIGYRPEMVNISFVGTDLALLAELEAG
AGVLSQVVPYFPDFIVFVYQAALALADPAAANGGVSFEGYLTGLRLAALAAAG
NPNTRSSYLTALQDLRDVLQQLQFGPGDQGLDQVLFELTISPAGTIAELTPAATP
VPAPCTG"
6897..9542
/ene="mcpA"
6897..9542
/ene="mcpA"
/codon_start=1
/transl_table=11
/product="methyl-accepting chemoreceptor"
/protein_id="AAC16114.1"

```

```

/db_xref="GI:3128262"
/translation="MROTMAISGATGDLAQAMVDILSAGTPCGDLQAAOHLQRTLAGL
DAALRDAPAGLREAVGTIGARAGDLVDARQOQFAALAEFDAVAGIFEVNLRISERLV
EIGDDAYENMWYMGGEAASGRVKTTLIEDLVDRDFARLSDALALRVVNVLRGAALWVP
GLDVAGQAIVRSDVAAGESRQDKIIVALEPTGLAPLRADLARLAGLARDLARPGRSD
DPQLQOSLQATKVDGLGSVDDLAFLALTNAEAGKANATTIDTLTRDVPAMIE
AARIEARADLVASALRLAALSLESYDRESAALAEARAVVAGOMAQPPDLRPLLRD
LIDRTDPARGLAQAHRLAKARAAEATFADAAAMDTITTCAAAAETVLGRIDGTS
GAVHRTSGAIGTLLLAGLSAIFGLLAPLAWLGLVPLRQVTOATLAAGDPFAV
DGLRPGCEIGALAGALTVPFRDALNDRPMREMDRAEAAAEAEAAQQAALAAAEV
ERDRDQQAQAEAEQRERRRURDAEAEAAQOQQLVSVRLAEALRAMSAGDLTA
RIETAFPAAYEGLRSDFNQAVTRMAELISEIVRASGSVSOEHAQSGSELGRRET
QAALAEETAAAMNQASVAQAALQCAAREARVQGTRETTAAGREVVRRITLQAMTDIA
QSSEQISRTSVIDDIAFQTNLLANAGVEARAGEAGRFVAVASEVGRSORDSEA
AREIAGLIATSVRQVGEVLAAGSDAALGRIEDLVLRDLALLAIAETGAGEOSAGIS
EITAVNOLDQVTOHNAAMFENNAATALLGEAGTCCSVRGLRQOGATGAAGAARL
TGRIRPSAQDYQDLALVRRGGHS"
9544..10413
/codon_start=1
/transl_table=11
/product="hypothetical protein"
/protein_id="AAC16115.1"
/db_xref="GI:3128263"
/translation="MQGRFDCGPALPGLLRGAGCKKGLMESRDPVVIHRELERYA
NAALKAGRIMESGASVSVHVGITMIARGFVETIGLSRGYTIAMTHVADQITIR
MLQVRLGNHRLDLAVRELCSRAERGMSEVEEAEARLVRPREPRAHAFWALTAVG
LACASFGLLGVDLAFQVWLAGAIGQYIRHLLHRGVNIFIVAGATAFLSATLGLL
GAIALQSTVOLAMMASILLVPGVPSTNAOTDINDGYTWSARAVVLMIMLFAAT
GIWFAELLGMR"
10416..10883
/codon_start=1
/transl_table=11
/product="hypothetical protein"
/protein_id="AAC16116.1"
/db_xref="GI:3128264"
/translation="MWDFLPHLLHQAELGAAAGFGVLNFGRLPWCRAAGALAL
MVRTGDLDAWSLEGATAAAIVTSCVTSALRGQGPACNVALAGCIPWVPAGAFGE
ALLGFPALTAPHPGNAVEIFVSAEHIIRVIFTTAAIGLAGLAIPAQLLRNODF"
complement(12010..12912)
/codon_start=1
/transl_table=11
/product="methylentetrahydrofolate dehydrogenase /
methenyltetrahydrofolate cyclohydrolase homolog"
/protein_id="AAC16117.1"
/db_xref="GI:3128265"
/translation="MADRRDHRIIDGKAVQORLLAGIRAELEAAAEHPVGRVLSISI
GPSPIEVYVRNQAPAAATAGLPFQVWWEASATQECACKRKIAALNDPDSVLGIILOR
PVPAHIVRSLOSAIHLKLDVEGMNPASIGNIVYNDIALAPCTAAAEVWVRATGLPL
TGPENVIGHSEIVGKPVAFLLMAEGATVTVCHHMTSRVAMHSRRADVIVAVGKAHL
IGPEMIFKPAVIDIGINQVTDAAQKVRILGDVDPTEAVREVAGWITTPVGGVPTVA
MLMRNALTARRRQVAAAGWLEAKA"
complement(13007..16723)
/ene="sbcC"
complement(13007..16723)
/ene="sbcC"
/codon_start=1
/transl_table=11
/product="potential exonuclease"
/protein_id="AAC16118.1"
/db_xref="GI:3128266"
/translation="MRILISQNTASLAAPFRIDFTTAPLAGAGLFAITGETGAGKS
SILDMACMLYGDAPRLSAGRSKVPDAAGEEISAADSRALRRGAATGWAERFAG
RDLEGYIARWQARRARDKVEKGLQTVARSLEAEQDQVLAQTOAQSOVQALGLSY
EEFRVTLAAGDQDFADRTGERATLEKVTGGLYRAVSVRYETENAREHAQ
LLAARAEHLDDARAAALTEIARATATAAABRAGLSEALARHRRHAEVRRQVQ
AAAALAGAAALQAEAGPARAQLRLEAPLHPWMAAARLARLARSAGAEAAA
SRSLAAQALATRADAAATAGTETEAFKAFGRINDRAAALDAQIATAATEAAR
SRAETAARAAQIRRAEDLAAEETRAQQAQAAEARLAEALAAQPLADDWPLRRDL
ADHRAACAAQAAATAAQADQALAQEAARAAAARQDAADSAALAKAAALQQA
LAPLEAAHPPTADLARIETDLAELTALRDGAFAAALGAQAQAPATAATEALAR
AQESAKQDLDRATQIHALTAPLEADLALSDAARSLRALSAGSCPVCVGCALHEPT
PAEAGLAHLAERLRADQAAAGAAQARDALTAAGARATAEARGTQAEODRRAQTR
AEAARAANGDTPQPRVSRPLALPGTDPDTLAAQAQDLRLCALQTAEEAAQAEISALR

```

CDS		<p> TRUTEARDRLRRALLAHRGTRERLAVQAAETAOEAAALAEARRTEAARRDGLALA LAPALRAGEDDPAAPGLAERLAATSAVGAARTGLOAAQAEASALAPQIAARRRSE TATAQAQAARADRDGAWAALRAERAPLLDQPTALHRSRNDORLQAAEQAAA ADLTAQAALAAAEARAAETARAASEAATQARAADLAAEAALAAEAAMSALAAALIA PPGTARLRLADLRGRDADVA"TAARSALGARKTDLOEI EDOGNPAEDPEALSARLAALIA EAPRAEIGCOLSAELARDAATRAALAGLAVETAAEAELVEAVNKATIGARSIGDRE ARIAQATLDVLVEHANRHLADLNPYRLRRAELSLQVEDRMDGTARATRSLSGGE RFVLSALALALVOMGKGTFAGTFLFIDEFGALDAGSLDOLAIDALETLOSQRQGV ISHVEAMQARIATRIAYRKEGGRRSLRVEGPGV" complement(16728..17945) /codon_start=1 /transl_table=11 /product="exonuclease SbcD homolog" /protein_id="AAC16119.1" /db_xref="GI:3128267" /translation="MRILHTADWHIGOTLNGWSREAEHRAFLADLGEILLAEQVDALL VAGVDFPLNPSGAEHLVLAALAGYVRANPRLQIVLTSGNHDPAPQLEAFLVRL GVHLGTLSPGSGMDLRHLPLRDRAGQIRAOVLALPFLROADLPGLRLGAEEGPE GAVTAPALLALAEIVARAALAGTLPILIAHMLTACAGLESAGNERILLIGDGHAVPP DVFPALAHVALHRLRPSLDGGRVYSGAPFLPSASEIGYDHGVTLTLLDLAGAAPP HIPLRPVPMRLRPAQATPLPEILAALDLRALDPRPAEGPFYLLALRADRPVTEIT AALDALEALPLRLAGLTITRPETARPAAPQDLTRTTPALFAAFLRLHGTEPET RHLSAERDALSEV" 18245..21130 /EC_number="1.6.5.3" /codon_start=1 /transl_table=11 /product="NADH-quinone oxidoreductase chain 5" /protein_id="AAC16120.1" /db_xref="GI:3128268" </p>	
CDS		<p> Query Match 6.4%; Score 262.2; DB 1; Length 189370; Best Local Similarity 52.8%; Pred. No. 6.9e-43; Matches 614; Conservative 0; Mismatches 543; Indels 6; Gaps 2; 526 CGGAGATTCCATGATTACGCGGAAACCCCTTAAGTCTCTCTGCGCAATGTCCAGGCTC 585 130052 CTGAATAATGACCCCTCTCGAACGGACCCCTTTCGGCCCTCCCGTCCCGCGCGC 129993 586 CCCCTATGATCAGACGGGATCAAGCTCGGATCGTGCATCTTCGGTGTAGGTAACTTTT 645 129992 CGCGTATGACGCGCGAGCTGACCCCGCGCATCTCATATCGGGCTTGCGCACTTTC 129933 646 TTGAGCCCATGAGCGGTTCTACGTCGAGCAGATCTTGAACAGCCTCCGACCTGGCGGA 705 129932 ACCGCGGCATCGCGGCTTTATCGGATGACCTTTTCGGCTTGGCGAGGGGACGACT 129873 706 TTGTTGGTGTGCGCTGACGGCGAGTGACCTTCAAGAAAAAGCCGAGGAATTCAGG 765 129872 GGGCGCTGATCGCGCGCGTGTCCGCGCCCGCGATGCCGCGATGCGTGAGTCTCTGG 129813 766 CCCGAGTCTGATTTCCCTGACCGAGCGGCTCCGTCGGGCAAGACAGGTCGCGC 825 129812 CGCAGGAGGCGTGTCTTCGTTGATCGAACTCGGCC--CGGGCGCCCATTTCCGCGCCG 129756 826 TCATCGGCGCGTGTGCTGACTATCTGCTTCCCGCGCGATCCGGAAGCCGTGTGAAGC 885 129755 TGATCGGCGGATGATCGACTTTGTCGGTGTGCAACCGCGCAATGCCCGCTGATCGCCA 129696 886 ATCTTGTGTATCCGGCCATCGCGATCGTGTTCATGACGATCAGGAAAGCGCGCTACAACA 945 129695 CGATGCGCGCGCGGAGATCCGATCGTCAGCTGACCGTACGAGGCGGCTATTACA 129636 946 TCAACGAGACGACCGGTGTTGATCTGGAGATCTGGGCGAGTAAAGCCGACCTCAAGA 1005 129635 TCGACCCGAAACACCGGACCTTCAGCCCGCATCACCGGACATTCAGCGCATGCGCGCC 129576 1006 ACCCGGAAAGCGCTACCGTTTTCGTTTACGTGTCGAGGCCCTCGTCTGCTGGG 1065 129575 GTCCCGATGCGGCTGACACCGGCTTTGGCGGATCATCCCGCGTGAAGATGCCCGCG 129516 1066 ATGCCGCGGTGAAGCATTTACGGTCTATGATTAACCTGATACCTGATCAACGGCAATG 1125 </p>	

Db	129515	CCGCGGCGGTGCGCCCTTTCACCGTGTATGTGTCGGAACAAGTCCCCCATACAGGCCATG	129456
Qy	1126	TCGCGCGCAAGGCTTCTCTCGGTATTCGGAAGCGCGCGATCCGGAGTTCGGCAAGTGA	1185
Db	129455	TCACCGCGATGGGTGGTGGCGGTCTTCGATCGGAACTGCGGACTGGA	129396
Qy	1186	TTCAGGAAACGCGACCTTCCCGAAACGGAATGTGTGATCGCATCACCCGACCGTTTCGG	1245
Db	129395	TCCATGCTCGGTGCGCTTCCCGAAATTCATGTCGACCGCATCACCCCGGCC---ACCG	129339
Qy	1246	CGGAATTCGCCAAGAAGCTCAACGCGGCCAGTGGGTGGATGACGACCTCCGCGTGTGG	1305
Db	129338	GCGATCGGAACGGCGCATGCGCGCTGATCGGCGCTTGACGATCGCGCGCGGTGACCT	129279
Qy	1306	CCGAGGATTTCCATCAGTGGGTGCTGGAAGACCAAGTTTCGGGATGGCGTCCGCGCTTG	1365
Db	129278	GCGAAACCTTTCGGGCAATGGGTGATCGAGGACAAATTCGCCGAGCGCGTCCGCGCTCG	129219
Qy	1366	AAAAAGCGCGGTGCAGATGGTGGGACGCTGACGAGCTGGGAGTACGTCAAGATCCGAA	1425
Db	129218	AAAAGTCCGCTGTCACCTTACCCCGCATGTCGACAAAGTTTCGAGACGATGAGATCCGCA	129159
Qy	1426	TGCTCAATGACGGGCTCATGCTCTCTTCCAGGCTTCCTGGTGGGTATGAGAAATG	1485
Db	129158	TCCTGAACGGCGGCGCATGCGCTATCCGCGCGGCTGGCGGACATCGCTTCG	129099
Qy	1486	TGGATGACGCCTTGAAGACAGCAACTCTTGGCATCTCAAGAACATATCTCAACAAGG	1545
Db	129098	TGCATGAGGCGGCGCGATCCGCGATCGCGCTTCTTGAACCGGCTGTCAGGCGCG	129039
Qy	1546	ATGTCATCCGACCCCTTGAAGCGCGCTTTCAGGCGATGACGCTCGAAGGCTATCGGACAGCG	1605
Db	129038	AGTGTGTCGCGATCGTCCCGCGGCGAGGATCTCAACGATTACAAAGCGCTGA	128979
Qy	1606	TCATCAGCCGCTTCTCAACAAGCGATGTCGGACACGAGCTCCGATCTCGATGCTAGCAGT	1665
Db	128978	TCATCAGCGGCTTCTCGAAACCCCGAAGTGGCGCGATGCGGTGCGCAGGCTTTCCTCGACG	128919
Qy	1666	GCTGTTCCAAAGTTTCAGTGTTC 1688	
Db	128918	GCTCGAACCGGACGCCGAAGTTC 128896	
RESULT 10			
AE009374			
LOCUS	AE009374	11164 bp	DNA linear BCT 20-DEC-2001
DEFINITION	Agrobacterium tumefaciens str. C58 linear chromosome, section 144 of 187 of the complete sequence.		
ACCESSION	AE009374	AE008689	
VERSION	AE009374.1	GI:17742928	
KEYWORDS	Agrobacterium tumefaciens str. C58 (U. Washington).		
SOURCE	Agrobacterium tumefaciens str. C58 (U. Washington)		
ORGANISM	Bacteria; Proteobacteria; alpha subdivision; Rhizobiaceae group; Rhizobium.		
REFERENCE	1 (bases 1 to 11164)		
AUTHORS	Wood, D.W., Setubal, J.C., Kaul, R., Monks, D., Chen, L., Wood, G.E., Chen, Y., Woo, L., Kitajima, J.P., Okura, V.K., Almeida Jr., N.F., Zhou, Y., Bovee Sr., D., Chapman, P., Clendenning, J., Deatherage, G., Gillet, W., Grant, C., Guenther, D., Kutyavin, T., Levy, R., Li, M., McClelland, E., Palmer, A., Raymond, C., Rouse, G., Saenphimmachak, C., Wu, Z., Gordon, D., Eisen, J.A., Paulsen, I., Karp, P., Romero, P., Zhang, S., Yoo, H., Tao, Y., Biddle, P., Jung, M., Kraspan, W., Perry, M., Gordon-Kamm, B., Liao, L., Kim, S., Hendrick, C., Zhao, Z., Dolan, M., Tingey, S.V., Tomb, J., Gordon, M.P., Olson, M.V. and Nester, E.W.		
TITLE	The genome of the natural genetic engineer Agrobacterium tumefaciens C58		
JOURNAL	Science 294 (5550), 2317-2323 (2001)		
MEDLINE	21608550		
PUBMED	11743193		
REFERENCE	2 (bases 1 to 11164)		
AUTHORS	Wood, D.W., Setubal, J.C., Kaul, R., Monks, D., Chen, L., Wood, G.E.,		

Chen, Y., Woo, L., Kitajima, J. P., Okura, V. K., Almeida Jr., N. F., Zhou, Y., Bovee Sr., D., Chapman, P., Clendenning, J., Deatherage, G., Gillet, W., Grant, C., Guenther, D., Kutyavin, T., Levy, R., Li, M., McClelland, E., Palmieri, A., Raymond, C., Rouse, G., Saenphimmachak, C., Wu, Z., Gordon, D., Eisen, J. A., Paulsen, I., Karp, P., Romero, P., Zhang, S., Yoo, H., Tao, Y., Biddle, P., Jung, M., Krespan, W., Perry, M., Gordon-Kamm, B., Liao, L., Kim, S., Hendrick, C., Zhao, Z., Dolan, M., Tingey, S. V., Tomb, J., Gordon, M. P., Olson, M. V. and Nester, E. W.

TITLE

JOURNAL Submitted (27-SEP-2001) Department of Microbiology, University of Washington, 1959 NE Pacific Ave, Box 357242, Seattle, WA 98195-7242, USA

FEATURES

source Location/Qualifiers

1. 11164
/organism="Agrobacterium tumefaciens str. C58 (U. Washington)"

/strain="C58"

/db_xref="taxon:180835"

94. 1578

/gene="mtlK"

/note="Atu4451"

94. 1578

/gene="mtlK"

/note="identified by sequence similarity; putative; ORF located using Blastx/Glimmer"

/codon_start=1

/transl_table=11

/product="mannitol 2-dehydrogenase"

/protein_id="AAL45245.1"

/db_xref="GI:17742929"

/translation="MTCKLSLATLDEAKTAAPAYSRNDLSAGIVHFGVGNFHRAHQ
AVYLLDFFNGADHFAIIGAGVLPDSATWREKLAQDLFTTVVEODNNRTGARVTGP
MIDLKVDGTDQIADITLADPKIRIVSMITTEGGVFIDAGSENPQHPAIAEDCKNPSA
PKTVGEFVAGLAKRDXDLQPTVMSCDNIPIHGKVTYKNAVVGCLAALSPPANWIG
ENAFPNMVDRIPTATGERERNIARDQFIEDNWPVCEEFKQWMEHNFPAGRPAL
EKAGQVQKVPAYPEHMKIRILNGHAAIAYPAALDHFVHEAMEHPLRFLAKLE
KDEIPIPVPTDLDNYFLERFLNPKIGDTIPRLAQDGSNRQPKFLPSTLDR
LSRGEDIVLSLWMCRYFYGTSDSGKEIVFNDAERLQAAAIKAKDDPVAFAL
DDIFGEVASALFRKRESHALTTLWQEGTAKTLQLYLDLDDKLAEK"

1585. 2304

/gene="Atu4452"

/note="Atu4452"

/note="identified by sequence similarity; putative; ORF located using Blastx/Glimmer"

/codon_start=1

/transl_table=11

/product="conserved hypothetical protein"

/protein_id="AAL45246.1"

/db_xref="GI:17742930"

/translation="MRDGTGPLLIFDCGVLDSEPVSIIVLLDMLSLHGVMTGEEEA
YERFLGRSVASMTTLFEEYGVETDIDFLDMRATLPERFTELKPIDGLAEYLDLLE
PLKRVASSQPERIRYSLGLTGLIDKFEPHVFSATVRNGKPAFDLFLHAARSMGVE
PROCVIEDSPAGIAAAGAGMGVFAFTGGSHARFPAPREKTAGLADGAVDAMPDLV
OLVGSYVGRGFGDQVERDAG"

2359. 3936

/gene="Atu4453"

/note="Atu4453"

/note="identified by sequence similarity; putative; ORF located using Blastx/Glimmer"

/codon_start=1

/transl_table=11

/product="ribitol kinase"

/protein_id="AAL45247.1"

/db_xref="GI:17742931"

/translation="MMRQNEVAVDVGTASARAGIFDPAGRLARSIHPILMQRPRENH
AEHSDTIWNAVCITAAALADAGVLPQSIATAGFDATCSLIVTDERGEPVSVSTDD
DRDPTIIVLDHRAIGEADRLTASGHRVLDFAAGNSVSPQMCKMLWLTHTMPOSWSRM
SEAFDLADFLTWATSAQSRNCTQTAKNFWLAQHPGWQADYLAFLAGLDDLKERAGL
PETVMPGAGIPLSPAAAEALGDTCCGVAAGMIDAYAGALGALGCGLAEDVGKHVA
LIATSSCLVAMSTQPMGRSLWPGPYWQAVLPGHVLEGGQSGATGALLDHIVRMHAAG

GEPTDTHARIVARVTELEGEFAFADRLHLVLPDFHGNRSPLADPHAVGVVSGLTLD
KSRIVDPATADAVLLGTAMTAATAGGVHVASLAAAGAAAMPGNAEISGNPALAAHYERD
YRFLAMRHRQOELES"

complement(4041. .4874)

/gene="Atu4454"

complement(4041. .4874)

/gene="Atu4454"

/note="identified by sequence similarity; putative; ORF located using Blastx/Glimmer"

/codon_start=1

/transl_table=11

/product="transcriptional regulator, Rpir family"

/protein_id="AAL45248.1"

/db_xref="GI:17742932"

/translation="MVEKLLNLTRAMKACTPSPERKTAKVLIHDELPEPTAOTLAA
KLSLSPMTVGRFLRSIGYRQFSDIRADLRHAEETVGADQPVTAEGGERHSNPFQSLLS
QQIOAIOTAFDMTTQTIWRLAMNETTSADDVCLATSPGDLGADRHFFGRULLECGHILH
YLGSDNAAVYALMDYDPARTLLVIMCGNLLPLQRLSTTARKGGYTLITTRFYEW
GPESVDLCAMPQSQNGGQILQILVLSLEFTLCALSAGADDAGKARVKNLAAALKRSLR
S"

4897. 5067

/gene="Atu4455"

4897. 5067

/gene="Atu4455"

/note="putative; ORF located using Glimmer"

/codon_start=1

/transl_table=11

/product="hypothetical protein"

/protein_id="AAL45249.1"

/db_xref="GI:17742933"

/translation="MNGDAFARPCCKRANHDLTVETSPPGHVLHRTGMRIIVDGLLYFR
AIVSRRIPAV"

5170. 5907

/gene="Atu4456"

5170. 5907

/gene="Atu4456"

/note="identified by sequence similarity; putative; ORF located using Blastx/Glimmer"

/codon_start=1

/transl_table=11

/product="conserved hypothetical protein"

/protein_id="AAL45250.1"

/db_xref="GI:17742934"

/translation="MIPDAARLAFGNLFAAETRSVFWKVLGTLTLLVLAALWFAIRSF
IYFALPMIDTLPICVPDAGNLSFVFAIFAGLIGLALALLISPTAVIAGLFLDDVA
DVVEKQNPDDPPGKAMPICEAVLSSIKFFGVTAGNLVALLLLLPVGNLIAPFLVN
GLLGRFFFEFAAMRFRSPADARLFRVKHRTTFVMAGLVIAAFIAPFLNLLTPFLFAA
SLMVHLHKAVSRRDPSPAGGTEQLRG"

complement(5869. .6579)

/gene="Atu4457"

complement(5869. .6579)

/gene="Atu4457"

/note="identified by sequence similarity; putative; ORF located using Blastx/Glimmer"

/codon_start=1

/transl_table=11

/product="endonuclease III"

/protein_id="AAL45251.1"

/db_xref="GI:17742935"

/translation="MRVKTAISKDELTEIFRRFSIORPEPKGELEHTNPFTLLVAVAL
SAQATDVGNRATRALFKVADTPKMLALGEEQLIGHIKTIGLYRNKAKNVIALSOML
IDNFGVEVPKTRRELVTLPGVGRKTANVMSMAFVPTLAVDTHVFRINRCLAPCK
TTDEVEDRLVRIIPEOYLFEHAHHWLIILHGRYCKCKARKEPCERCVCVIADICKSPEKTFDI
PAPLVELPAQLFGAAGGE"

6689. 7171

/gene="Atu4458"

6689. 7171

/gene="Atu4458"

/note="identified by sequence similarity; putative; ORF located using Blastx/Glimmer"

/codon_start=1

/transl_table=11

/product="conserved hypothetical protein"
/protein_id="AAL45252.1"
/db_xref="GI:17742936"
/translation="MDLNDOPFAAELVPHRSRGRFRLLILLTGFACLLYGGFEL
ATGWPICLFFGLDFLLIYAFNRAAKAREVSVSRNLSIRKSPAGRWVEYFR
NPFWRFRVRHDEIGIVSMHVITGEGRATDIGSEFLNPDDESFAKAFGGALATVRRRM
".

gene

7438..8310

/gene="ada"

/note="Atu4459"

7438..8310

/gene="ada"

/note="identified by sequence similarity; putative; ORF
located using Blastx/Glimmer"

/codon_start=1

/transl_table=11

/product="6-O-methylguanine-DNA methyltransferase"

/protein_id="AAL45253.1"

/db_xref="GI:17742937"

/translation="MNANIMLNEDITPIGSDYDVRVGVIELLLTDYREQPSLEAIAAR
LGQSPLOKTFTRWAGLSKPAFLQAVTLDHAKLLREEDLPLLETSTIEVMSGPGRL
HDLVTHENSPEWKAAGGLTIRYGFHASPFGLALVIMTDRGLASCAADPDGERA
CFEDMAGRNADIVREATAPIARIFEPAMWTADKPLRVVLGIDFDQVRWKSLL
KIPMRVATYSNIACDIGOPTASRAVGAAGVNPVSVFVPCHRVAVGSGALTYHWGL
TKRAMLGWETGKA"
8375..8599

gene

Query Match 6.3%; Score 259.2; DB 1; Length 11164;

Best Local Similarity 54.3%; Pred. No. 3.8e-42; Indels 9; Gaps 3;
Matches 591; Conservative 0; Mismatches 486;

Qy	581	GGCTCCCCCTATGACATCGACGGATCAAGCCCTGGATCGTGGATTTTCGGTGTAGTAA	640
Db	147	GGTTCGGGCTATTCCCGGAAGATCTGTCGCGCGCATCGTGTATTTCGGCTTGCAA	206
Qy	641	CTTTTTCGACCATGAGCGGTCTACGTGACGAGATTCTTGAACACGCTCCGGAGTG	700
Db	207	TTTCACCGCCCATCAGCGGGTCTATCTTGTGATGACCTGTCAACACCGGTGGGACCA	266
Qy	701	GGCGATTCTTGTGTGGCTGACGGGCGAGTGACGGTTCAAGAAAAAGCCGAGGATT	760
Db	267	CGACTTCCCATCATCGCGCGCGGGTCTCCCTCCGATCCGAGATGCCGGAAGAACT	326
Qy	761	CAAGCCCGACGATGCTGTATTCCCTGACCGACAGCGCTCCGCGGCAAGACACGGT	820
Db	327	TGCCACACAGATTTCCTGACCACTGTCTGCGAGCAGCAATAATCGACCGGTGC---	383
Qy	821	CGCGTCATGGCGGCGTCTGCTACTATCTGCTTCCCGCGCGATCCGGAAGCGGTCT	880
Db	384	CGCGTCACCGGCGCGATGATCGACATCTCGAAAGT---TGGCGATACCCAGGCGCAT	440
Qy	881	GAAGCATCTTGTGATCGGCGCATCCGATCGTTTCCATGACATCAGGAAAGCGGCTA	940
Db	441	CGATACGCTTCCGATCCGAAATCCGATCGTTTCGATGACGATTACGAGGCGGTTA	500
Qy	941	CAACATCAACGACGACCGGTGGTTCGATCTGGAGAAATCGGCAGTAAAGGCGCACT	1000
Db	501	TTTTCAT---CGATGCTCCGGCTCTTCAATCCGACGATCCGGCATTCGCCGAAGCG	557
Qy	1001	CAAGAACCCGGAAGCGGTACCGTTTCGGTTCAGTGTGTCGAGGCCCTTCGCTCGC	1060
Db	558	CAAGAACCCGTCGCTCCGAAACCGGTTCGGTTTCATCGTTCGCGGGCTGAAGCGCG	617
Qy	1061	TTGGATCGCGGTGTGAAGCATTTACGGTCATGTCCTGTGATAACCTGCGCTATAACGG	1120
Db	618	CGCGATGAAGACCTCGACCGCTTACCGTCTATGCTCTCGACACACATTCGCCACATGG	677
Qy	1121	CAATGTCCCGCAAGCGCTTCTCGGCTATGCCAAGCGCGCGCATCGGAGTTGGCGAA	1180
Db	678	CAAGGTGACGAAAAACCGGTTGGTTCGGCTGCGAGCGCTCTCCGACCGCGCTTTGCCAA	737
Qy	1181	GTGATTGAGAAACCGGACCTTCCCGAAGCGGAATGTTGATCGCATCACCCCGACCGT	1240

```

/product="AGR_L_810p"
/protein_id="AAK88977.1"
/db_xref="GI:15158761"
/translation="MTDDSSERYRAPALDKGLDILELLIARLTDGGLTOLEIAKATGKSP
NELVRLDLRVGRVYVLTQDGRFSLTLKMFGLAHFAPIRRLVFAAPVMRDPFSASA
EOACHLAVDRCSVVVTAQOESATYWMKSMRVGAQMSLFHTGSGHVILLAFQTEAOREI
MPCQVGRSGTVPADIAERURQVRNMGFSFMSDSLOTAGVRNISAPILTLDDGTALAVI
TCTPITLGGTAPREFCEQILRDAAKRISEVVTGNV"
complement(972..1937)
/gene="AGR_L_812"
CDS
/gene="AGR_L_812"
complement(972..1937)
/note="transaldolase PA2796 {imported} - Pseudomonas
aeruginosa (strain PA01)"
/codon_start=1
/transl_table=11
/product="AGR_L_812p"
/protein_id="AAK88978.1"
/db_xref="GI:15158762"
/translation="MTSKLEQLRAITTVADTGDIQAVARLKPVDCTNPTIVLKAIG
TPAFADAVKEAVANGRGQSDAVAAVADRLAISVGAALGLVPGRVSTEVVDALIS
PDQASITKARHITASYKRGIERILIKLASTWEGIKAAEVLOSIGIDCNLTLLFS
QAOAIAACAERAKAFILISPPVGRILDMYKSTGETYATETDPCGVSVRSIYXYKANGIG
TYVGCASFRRNVGEALAGCDRLTISPALLEELKDNGKLERKLSPNVRAEALQSLD
EKAFRWAMNEMATEKLSEGLRFAKDLVTLREMVVRKELTAAAA"
2133..3164
/gene="AGR_L_814"
CDS
/gene="AGR_L_814"
2133..3164
/note="(AP001509) transcriptional regulator"
/codon_start=1
/transl_table=11
/product="AGR_L_814p"
/protein_id="AAK88979.1"
/db_xref="GI:15158763"
/translation="MAKTAKLRRETHITYSEASLSRLRAWLXYNGMTCQKDVAEEL
GISRSTVIRLLDEAMKSEVQIWINIEGIDFVSLAQLEAYGLDEAVITIPSPGPSA
RPSTCVTSKTSKSAETGAKAVGALGOLFSEVVPNGATIGVWGRTMTASLSFRPP
GRNCKVYSLGGTVAVHOTNPLDTWRLASALGAECYMFELAPLLVSFETKRALIEK
CGLATLDLAETLADLVSCDGIIPHSTISLSEGIISRETRELVDAGCVDCTMFNFID
AEGRSVDHPINORAMAIDLDTLRKAKHIVLASGGAHRAIAIRATIKRIGCWTLTDEA
AARALMELV"
complement(3161..3448)
/gene="AGR_L_815"
CDS
complement(3161..3448)
/gene="AGR_L_815"
/note="HYPOTHETICAL 11.0 KD PROTEIN Y4KP"
/codon_start=1
/transl_table=11
/product="AGR_L_815p"
/protein_id="AAK88980.1"
/db_xref="GI:15158764"
/translation="MRLIWTFRYRELDAIGVYIAERNPRAAAKVVRAIHOTTARLLS
DNPHLGRSGEGTRELVPGLPIYAVRYTEEOVDIVFVQHAAREWPDDA"
complement(3454..3690)
/gene="AGR_L_816"
CDS
complement(3454..3690)
/gene="AGR_L_816"
/note="hypothetical protein"
/codon_start=1
/transl_table=11
/product="AGR_L_816p"
/protein_id="AAK88981.1"
/db_xref="GI:15158765"
/translation="MKSTIELSDDIKRRDLTAPRSNSTPSRIIEDALSHGSRSLAWQE
KWTSGVRAGLAEEADAGEVTEEEINDVKNYKA"
complement(3743..4615)
/gene="AGR_L_818"
CDS
complement(3743..4615)
/gene="AGR_L_818"
/note="O6-methylguanine-DNA methyltransferase PA2118
(imported) - Pseudomonas aeruginosa (strain PA01)"
/codon_start=1

```

```

/transl_table=11
/product="AGR_L_818p"
/protein_id="AAK88982.1"
/db_xref="GI:15158766"
/translation="MNANIMLNEDIPTIGSDYDTRVGVIELLTLDYREOPSLAIAAR
LGQSPTLOLQKFTTWAGLSKPAFLQAVTLDHAKRLLEDPLLETSTEVGMSGPGRLL
HDLFTHEAMSPGEMKAGKGGLTIRYGFHASPFGALVMTIDRLAGACAFADPDGERL
CFDMAGRNPNADVEDREATPYAARIFEPMATMTADKPLRVVLGLTDFQVRYVKSLL
KIPMGRAVTVYSNIACDICOPTASRAVGAAGNPVSFVVPCHRAVGKSGALTYGHWGL
TKRAMLGWETGKA"
complement(4882..5364)
/gene="AGR_L_819"
CDS
complement(4882..5364)
/gene="AGR_L_819"
/note="hypothetical protein"
/codon_start=1
/transl_table=11
/product="AGR_L_819p"
/protein_id="AAK88983.1"
/db_xref="GI:15158767"
/translation="MDTLNDOPFAAELVPHRSRGRGFRLLILITGTFACLIYGGFFL
ATGAMPICLFGDLFLLYIAFRANYRAAKAREVSVSRNLNLSIRKFSAGRMVYERF
NPFWARFVRRHDEIGIVSMHVTGEGRATDIGSLNDDRESFAKATGGALATVRRRM
"
5402..6184
/gene="AGR_L_820"
CDS
5402..6184
/gene="AGR_L_820"
/note="endonuclease III PA3495 {imported} - Pseudomonas
aeruginosa (strain PA01)"
/codon_start=1
/transl_table=11
/product="AGR_L_820p"
/protein_id="AAK88984.1"
/db_xref="GI:15158768"
/translation="MTVAKKRSSTQTLKKSIPAORRPPVRVKTAYSKDELTEIFRRFS
IQRPPEGIEHTNPTFTLLVAALSAQATDGVNRRATRAKFKVADTPEKMLALGEEQL
IGHKTLGLVRYNKAKNVIALSOMLIDNFGVEGPKTRELVTLPVGRKTKANVMSMAF
LPVLAIVTHVRIANRCLAPGKTDEVEDRLVRIIPEOYLPHAHHLWILLHGRYCK
ARKPECRCVIADICKSPEKTFDIPALVELPAQLFGAAGE"
complement(6146..6883)
/gene="AGR_L_822"
CDS
complement(6146..6883)
/gene="AGR_L_822"
/note="hypothetical protein"
/codon_start=1
/transl_table=11
/product="AGR_L_822p"
/protein_id="AAK88985.1"
/db_xref="GI:15158769"
/translation="MIFDAARLAFGNLFAAETRSVFWKVLGLTLLVLAALWFARSIF
IYFALPWIDTLIPGVDPWAGWLSFVFAIFAGIGLALALLISPTAVIAGLFLDQDVA
DVEPKONYDDPGKAMPICEAVLSSIKFEGVVIAGNLVALLLLVPCVNLIAFELVN
GYLLGRFFEFEAAMRFRSPADARLFRVVKHRTTVFMAGLVIAAFVLAVFLNLLTLPFAA
SLMVHLHKAVSRDRDPSPAAGTQOLRG"
7179..8012
/gene="AGR_L_824"
CDS
7179..8012
/gene="AGR_L_824"
/note="hypothetical protein"
/codon_start=1
/transl_table=11
/product="AGR_L_824p"
/protein_id="AAK88986.1"
/db_xref="GI:15158770"
/translation="MVEKLLNLNUTRAMKAGTSPERKIAKYLIEHLDLPETTAOTLAA
KLISLSPMTVGRFLRSLGRFQFSDIRADLRHAEETVGADQPVTAEGGERHSPFQSLLS
QIOAIOIQTAFDMTQTITWRLANNETSADDDVCLATSPGLDAGRHFGYRLLCECRGHIH
YLGSDAAVYVALWDYDITLIMDCGNLLPLQRLSTTARKGGRYTLITTTFFYEW
GPESVDLCAMPQSQNGQGLQLQVLSLEFTLCALSAGADDAGKARVKVNLAAALKRSLR
S"
complement(8117..9694)
/gene="AGR_L_826"

```

CDS	complement(8117..9694) /gene="AGR_L_826" /note="D-ribulokinase (EC 2.7.1.47) - Klebsiella pneumoniae" /transl_table=1 /codon_start=1 /product="AGR_L_826p" /protein_id="AAK88987.1" /db_xref="GI:15158771" /translation="MMRNLVAVDVGTASARAGIFDPAGRLIARSIHPILMQRPRENH AEHSDIWAACVLAADALADAVLPSTAAIGFDATCSLVIRDERGEPVSSTDD DFDTIYWLHRAITGEADRLTASGRHVSPEMQMQLMWLTKTHMPSWSRM SFADLADFTWKGATQASRNSCTQTAKWFLAQHPGQADYLAFAGLDLDRKRAGL PETVMPGGATGLPSAAQAEGLDGTQCVAAAGMIDAYAGALGALGCAEDVGVKHA LIAGTSSCLVAMSTPGPSLWGPVQWALPGHVLVEGGSATGALLDHTVRMHAAG GEPTALHVARIVRTALRELEGEAFADRLHVLDPFHGNSPLADPHAVGVVSLGLTD TSFSDSLRLYARVTAIALGARNHVLDAHERGYAVETLHVGTGHVKNPLLMELYADVT GKRIVFATDAVLLGTAMTAANTAGGVHASLAAGAAAMPNGNAEISNPALAAHYERD
Query Match	6.3%; Score 259.2; DB 1; Length 14945;
Best Local Similarity	54.3%; Pred. No. 3.7e-42;
Matches	591; Conservative 0; Mismatches 488; Indels 9; Gaps 3;
QY	581 GGCTCCCCCTATGACATCGACGGGATCAACGCTGGGATCTGTCGATTTCGGTGTAGGTAA 640
Db	11906 GGTTCGGGCTATTCCCGGAACGATCTGTCCGCGGCGATCTGTCGCGGTTCGGCA 11847
QY	641 CTTTTTCGACCCATGAGCGTCTAGCTGAGCAGATCTTGAACACGCTCGGACTG 700
Db	11846 TTTTACCGCCCATCAGCGGCTCTATCTTGATGACCTGTTCACACCGGTGCGGACCA 11787
QY	701 GCGGATTGTGGTGTGGCCCTGACGGGCGAGTGACCGCTTCAAGAAAAAGCCGAGGAATT 760
Db	11786 CGACTTCGCCATCATCGCGCGCGGCTTCTGCCCTTCGATGCCAGTCCGCGGAAACT 11727
QY	761 CAAGCCCAAGACTGCTGTATTCCTGACCGAGACGGCTCCGTCGGCGACAGACGACG 820
Db	11726 TGCCGCACAGGATTTCCTGACCACTGTCTGTCGACGAGCAGACAATAATCGACCGTGC --- 11670
QY	821 CGCGTTCATGGCGCGCTGCTGACTATCTCTTGCCTGCCCCCGCATCGGAAGCGGTCT 880
Db	11669 CGCGTACCGGGCGGATGATCGACATCTCAAGT---TGGCATACCCAGGCCATCAT 11613
QY	881 GAAGCATCTTGTGATCGCGCCATCCGATCGTTTCCATGACATACGGAAGCGGCTA 940
Db	11612 CGATACGCTTGCATCGGAAATCCGATCGTTTCGATGACGATTACCGAGGCGGTTA 11553
QY	941 CAACATCAACAGACGACCGTGTTCGATCTGGAGATCGGCAGTAAAGGCGGACCT 1000
Db	11552 TTTTCAT---CGATGCTCCGGCTCTTCAATCCGACGATCCGGCCATTCGCGAAGACGG 11496
QY	1001 CAAGAACCCGAAAGCGCTACCGTTTTCGGTTACGTGTGCGAGGCGCTGCGTCGTCG 1060
Db	11495 CAAGAACCGTCCGCTCGAAACCGTGTTCGGTTTCATGTCGTCGGGCTGAGGCGCG 11436
QY	1061 TTGGATCGCGGTGTAAGGATTTACGGTTCATGCTCTGATACCTGCTGATACAGG 1120
Db	11435 GCGGATAAGACCTGACGCGCTTACCGTTCATGCTGCGACACATTCGCCCAATGG 11376
QY	1121 CANTGTCGCCGACGACCTTCTCGGTATCGAAGCGCGCGCATCCGGAGTTGGCGAA 1180
Db	11375 CAAGGTGACAAAACCGGCTGCGGCTGCGAGCGCTCTCCGACCCGCGCTTGGCCAA 11316
QY	1181 GTGGATTCAGAAAACGCGACCTTCCCGAAGGATGTTTCATCGCATCACCCCGACCGT 1240
Db	11315 CTGNATCGCGGAAATATGTCGCTTCCGAATTCATGTCGACCGCATCACCCCGCGAC 11256
QY	1241 TTCGCGGAAATCCCAAGAGCTCAACGCGCGCGCATGGGTGGATGACGACCTGCCGCT 1300
Db	11255 CGGCGAGGCTGAACGCAATATCGCCGCTGACCATTTTCGGCATCGAGGATAATTGCGCGT 11196
QY	1301 GGTGCCGAGGATTTCCATCATAGTGGTCTCGAAGACCATGTTTGGGATGCGCCGCC 1360
Db	97474256

Db	11195 CTTCTGCGAAGAGTTCAAGCAATGGTGTATGGAGGATAATTTCCCGCAGCGCTCGGC 11136
QY	1361 GCTTGAAGAAACCGCGCTGCAGATGGTCGGGAGCTGACGAGCTGGGAGTACGTCAAGAT 1420
Db	11135 GCTGAGAAAGCGCGCTGCAGTTCCGTAAGAGTGTTCGGCCCTTACGAGCACATGAAGAT 11076
QY	1421 CCGAATGCTCAATGACGGGATGTCTGCTCTCCAGGCATTTCTGGTCGGCTATGA 1480
Db	11075 CCGCATCTGAATGGTGGCCACCGGCANTCGCTATCCGCGACGCTGCTCGACATCA 11016
QY	1481 GAATGCTGATGACGCCATTTGAAGACAGGAACTCCTTGGCAATCTGAAGAACTATCTCAA 1540
Db	11015 TTTGTCGATGACGGGATGGAGCATCCGCTGATCCGGCGGTTTCTGGCCAAAGCTCGAAA 10956
QY	1541 CAAGGATGTATCCCGACCCCTGAAGGCGCTTCAAGGATGACGCTCGAAGGCTATCGGA 1600
Db	10955 AGACGATCATCCCGTTCATTCGCGGTACCTTGACACCGACCTCAACGACTATTTCCG 10896
QY	1601 CAGCGTTCATCAGCGTTTCTCCAAAGGCGATCTCGGACAGACGCTCCGGATTTGCTAG 1660
Db	10895 GCTCATCGAGCGCGCTTCTCTCAACCCGAAGATTGGCGATACCATTCGCGCCTCGCGCA 10836
QY	1661 CGATGGCT 1668
Db	10835 GGATGGTT 10828

RESULT 12	AF018073	9810 bp	DNA	linear	BCT 22-OCT-1997
LOCUS	AF018073				
DEFINITION	Rhodobacter sphaeroides operon regulator (smoc), periplasmic sorbitol-binding protein (smoE), sorbitol/mannitol transport inner membrane protein (smoF), sorbitol/mannitol transport inner membrane protein (smoG), sorbitol/mannitol transport ATP-binding transport protein (smoK), sorbitol dehydrogenase (smoS), mannitol dehydrogenase (mtlK), and periplasmic mannitol-binding protein (smoM) genes, complete cds.				
ACCESSION	AF018073	L13697			
VERSION	AF018073.1	GI:2338757			
KEYWORDS	Rhodobacter sphaeroides.				
SOURCE	Rhodobacter sphaeroides.				
ORGANISM	Bacteria; Proteobacteria; alpha subdivision; Rhodobacter group; Rhodobacter.				
REFERENCE	1 (bases 6163 to 8431)				
AUTHORS	Schneider,K.H., Giffhorn,F. and Kaplan,S.				
TITLE	Cloning, nucleotide sequence and characterization of the mannitol dehydrogenase gene from Rhodobacter sphaeroides				
JOURNAL	J. Gen. Microbiol. 139 (Pt 10), 2475-2484 (1993)				
MEDLINE	94075966				
PUBMED	8254318				
REFERENCE	2 (bases 5865 to 6635)				
AUTHORS	Schneider,K.H. and Giffhorn,F.				
TITLE	Polyol metabolism of Rhodobacter sphaeroides: biochemical characterization of a short-chain sorbitol dehydrogenase				
JOURNAL	Microbiology 141 (Pt 8), 1857-1863 (1995)				
MEDLINE	96032011				
PUBMED	7551049				
REFERENCE	3 (bases 6679 to 8212)				
AUTHORS	Schaefer,A., Stein,M.A., Schneider,K.H. and Giffhorn,F.				
TITLE	Mannitol dehydrogenase from Rhodobacter sphaeroides S14: subcloning, overexpression in Escherichia coli and characterization of the recombinant enzyme				
JOURNAL	Appl. Microbiol. Biotechnol. 48 (1), 47-52 (1997)				
MEDLINE	97419532				
PUBMED	9274047				
REFERENCE	4 (bases 4855 to 6750)				
AUTHORS	Stein,M.A., Schaefer,A. and Giffhorn,F.				
TITLE	Cloning, nucleotide sequence, and overexpression of smoS, a component of a novel operon encoding an ABC transporter and polyol dehydrogenases of Rhodobacter sphaeroides S14				
JOURNAL	J. Bacteriol. 179 (20), 6335-6340 (1997)				
MEDLINE	97474256				

9335280
 PUBMED
 REFERENCE
 5 (bases 6163 to 8431)
 AUTHOR
 Schneider,K.H. and Giffhorn,F.
 TITLE
 Direct Submission
 JOURNAL
 Submitted (09-MAR-1994) Lehrstuhl fuer Angewandte Mikrobiologie,
 Universitaet des Saarlandes, Geb. 2, Saarbruecken D-66041, Germany
 REFERENCE
 6 (bases 4855 to 6162)
 AUTHOR
 Stein,M. and Giffhorn,F.
 TITLE
 Direct Submission
 JOURNAL
 Submitted (22-MAR-1996) Lehrstuhl fuer Angewandte Mikrobiologie,
 Universitaet des Saarlandes, Geb. 2, Saarbruecken D-66041, Germany
 REFERENCE
 7 (bases 1 to 9810)
 AUTHOR
 Stein,M. and Giffhorn,F.
 TITLE
 Direct Submission
 JOURNAL
 Submitted (11-AUG-1997) Lehrstuhl fuer Angewandte Mikrobiologie,
 Universitaet des Saarlandes, Geb. 2, Saarbruecken D-66041, Germany
 COMMENT
 On Aug 21, 1997 this sequence version replaced gi.1236173.
 FEATURES
 Location/Qualifiers
 1..9810
 /organism="Rhodobacter sphaeroides"
 /strain="Si-4"
 /db_xref="taxon:1063"
 681..1634
 /gene="smoc"
 681..1634
 /gene="smoc"
 /note="ABC transporter; putative; Smoc"
 /codon_start=1
 /transl_table=11
 /product="operon regulator"
 /protein_id="AAC45765.1"
 /db_xref="GI:2338758"
 /translation="MIDRPESEPTPLDARAGWLYVAGLTQDQIARELQTSRQAO
 RLVSRAIERLHVRHVSQCLHLEAALLRRFGLKLARVAPSLGSEVDPLPSIAPT
 AAEEVRLRERPMVAFGTGRSLRATVEEMTSMVCQHKIVSLNGNISADGSASY
 DVIFRIARVAPHVPMPIVTAQDAARELFHALKPVQSVLRARNADYFVGLGQM
 GEDAPLLKDGITPELTEMQDLGAVAGVDFDESGRYLETSINQKAVRVRELSE
 DRTVVAIGAGRRKLAALHAGLRGLNFLITDEFTAQALLS"
 1663..3052
 /gene="smoe"
 1663..1668
 /gene="smoe"
 1687..1692
 /gene="smoe"
 1731..1735
 /gene="smoe"
 1742..3052
 /gene="smoe"
 /note="ABC transporter; putative; Smoe"
 /codon_start=1
 /transl_table=11
 /product="periplasmic sorbitol-binding protein"
 /protein_id="AAC45766.1"
 /db_xref="GI:2338759"
 /translation="MTARFRALMGACAAVALSSRAGAEITIVATVNVNGDMTRMOGLMS
 EFNQAPDITVEWTLLENLRQKVTITDIATKGGOFDVLTITGYEVPWKGQGLVSL
 NDLPEDDADILLPAIRNGLTVDGELYAAPFGSSIMIMYRKDLMEKAGLTFMPDAPT
 DFVKAAGKTDKDAEYVYICLRKAGKGNNAFLSAMANSYGARWFENQPDGGE
 AKWATLDYLDWMYNGPPGKNGFNENLALFOQKCGKMWIDATVAASFVTPNEEST
 VADKVGFLAPDTCGKFPANLGAWNLAIPAGSKQKVDAAKQFIATATSKDYAEVLASK
 EGWAVPGRITSLYENPEYKVPFPAKMTLSINRAADTPHFAVDPPVYGVGVVAIPE
 FCGTGTAVGQGFSAALGMSAEALQAQAOFTTREMTRAGYIK".
 3061..3088
 /note="smoe-smoe intergenic region repeat"
 3102..3120
 /note="smoe-smoe intergenic region repeat"
 3144..4025
 /gene="smof"
 3144..3147
 /gene="smof"
 /note="putative"
 3153..4025
 /gene="smof"
 repeat_region
 3061..3088
 /note="smoe-smoe intergenic region repeat"
 repeat_region
 3102..3120
 /note="smof-smof intergenic region repeat"
 gene
 3144..4025
 /gene="smof"
 RBS
 3144..3147
 /gene="smof"
 /note="putative"
 CDS
 3153..4025
 /gene="smof"
 /note="ABC transporter; putative; Smof"
 /codon_start=1
 /transl_table=11
 /product="sorbitol/mannitol transport inner membrane
 protein"
 /protein_id="AAC45767.1"
 /db_xref="GI:2338760"
 /translation="MATQHSKTAARLMISPAVILLFLMWIVPLSMTLTYSFSLRYNLLM
 PGMSEFTGMDNYYFLTPDAFSAALTNTILLVVGVLITVVGVLALLLLDQPPFMGQG
 IVRVLVIAPEFVMTVSALVKNMFNMVNGMFEAHIARGLGLPPDFLSQAPLASIIG
 IVANQWLFPAILILLTALQSLDREOMEAEMDGASALDRFHIHTVPHLITRAITVVVLI
 QTIFLLGVFAEILVTNGGPGTASTNITLYLVAOSLLNVDVGGSGAGGIVAVVLANIV
 AIFLMRMIGKNLDA"
 4016..4857
 /gene="smog"
 4016..4020
 /gene="smog"
 /note="putative"
 4027..4857
 /gene="smog"
 /note="ABC transporter; putative; Smog"
 /codon_start=1
 /transl_table=11
 /product="sorbitol/mannitol transport inner membrane
 protein"
 /protein_id="AAC45768.1"
 /db_xref="GI:2338761"
 /translation="MSRRTSTRRTLIIVTLAAWTIAFLIFFPILMTVLMSPKSEGDALK
 APPAMFLSDWTQSYADVOERSNVARHEMNSVVISLSTLVALAIAIPAANAMAEVPG
 RTKADVLMMLSLTKMMPAVGLIPLIYLRFDRTGLDTRIGLIVLTLNLPITVVMVLY
 TYFKEIPGEILEAARMGDATLGSEILYILTPMVGPIASTLLNLLNVAWNAFAFWTLQI
 TTSRAAPLTQPIASYSSPEGLFYAKLSAATMAIAPILILGWFSOKOLVRGLTFGAVK
 ".
 4857..5863
 /gene="smok"
 4857..4860
 /gene="smok"
 /note="putative"
 4865..5863
 /gene="smok"
 /note="ABC transporter; putative; formerly polk; Smok"
 /codon_start=1
 /transl_table=11
 /product="sorbitol/mannitol transport ATP-binding
 transport protein"
 /protein_id="AAC45769.1"
 /db_xref="GI:2338762"
 /translation="MGKITLRNVOKRFGAEVVIPLSLDIEDGEEVFVFGSPGCGKST
 LLRLTAGLEDVSDGIMIDGRDAYTEMPPAKRLAMVFQSYALYPHMTYKKNIAFPPLRM
 AKMEPQETERRVSNAAKILNLTNYLDRRPGQSGQQRQVRAIGRAIVREPAALFDEP
 LSNLDAALRVNMRLEITELHOSLETTMIYTHDOVEAMTADKIVLNAEIRIEOVGSP
 LTLVRNPANLFAVGFIGSPKMLLEGPEAAKHGATTIGIRPEHIDLREACAGWGEVGV
 VSEHLGSDTFLHVVHAGNPTLTIVRTGEGFVGHGDRVWLTQADKILHRFGADGAL".
 5852..6635
 /gene="smos"
 5852..5859
 /gene="smos"
 /note="putative"
 5865..6635
 /gene="smos"
 /EC_number="1.1.1.14"
 /note="short-chain alcohol dehydrogenase; formerly pois;
 Smos"
 /codon_start=1
 /transl_table=11
 /product="sorbitol dehydrogenase"
 /protein_id="AAC45770.1"
 /db_xref="GI:2338763"
 /translation="WRLDGKTLITGSARGIGRAFAEAYVREGARVALADINLEAARA
 TAEIGPAACAIALDVTDOASIDRCVAELDRWGSIDILVNNALFDLAPIVEITRES
 YDRLFAINVSGTLFWMQAVARAMIAGRGKKLIINWASQGRGALVGVYCATRAAVI
 SLTQSAGLNLIRHGINIAPVVDGHEWDGDAKFDYENLPRGKKRQVGAAPF
 GRMGRAEDLTGMAIFLATPEADYIVAQTYNVNDGGNNNS"

```
gene 6680..8124
/ gene="mtik"
RBS 6680..6684
/ gene="mtik"
/ note="putative"
CDS 6691..8124
/ gene="mtik"
/ EC_number="1.1.1.67"
/ note="mtik"
/ codon_start=1

Query Match 6.0%; Score 247.4; DB 1; Length 9810;
Best Local Similarity 53.4%; Pred. No. 9.4e-40;
Matches 590; Conservative 0; Mismatches 506; Indels 9; Gaps 3;

QY 584 TCCCCCTATGACATGACGGGATCAAGCCCTGGGATCGTGCATTTCCGTTAGTAACTT 643
|||||
DB 6711 TCCCTCTATGACCGCAAGCGCTCACTCCGGGATCGTCCATATCGCGGTCGCAACTT 6770

QY 644 TTTTCGAGCCCATGAGCGGTTCTACGTCGAGCAGATCTTGAACACGCTCCGGACTGGC 703
|||||
DB 6771 CCACGGGGGATGATGCGGCTCTATCTCGAGATCTTCTCGCGTGGCGAGGCCACGA 6830

QY 704 GATTGTTGGTGTTCGCTGACGGCAGTGACCGTTCAAAGAAAAAGCCGAGGAATTCAA 763
|||||
DB 6831 CTGGGCCATCTCGCGCGCGCGCTCCGCCACCGATGCGCGGATGCGCGAGGCTTGGC 6890

QY 764 GCGCCAGGACTGCTGATATCCCTGACCGAGACGGCTCCGTCGGCAGACGACGGTGGC 823
|||||
DB 6891 CGCGAGGACAAATCTCTGACGGGTGAGCTCGATCGCGGGGCCA---CCGGGGCCG 6947

QY 824 CGTCATGCGCGCTGCTGACTATCTGCTGCCCGCCGATCGGAAGCCGTCTGAA 883
|||||
DB 6948 CAGGTGGGGGATGGTGGCTTCTCGCGTGCAGGCCACAAATCGGCGCTGATCGA 7007

QY 884 GCATCTTGTGATCCGGCCATCCGATCGTTTCATGACATCATCAGGAAGCGGCTACAA 943
|||||
DB 7008 GGCCATGTCGGATCGCGCATCCGATCGTCTGCTGACCGTACCGAGGCGGCTATTA 7067

QY 944 CATCAACGAGACGACGGTGTCTGATCTCGAGATCGGACATCGGACGTAAGACCGACCTCAA 1003
|||||
DB 7068 TGTGATGCTCG---GGCGCTTCGATCCGACGATCCCGATATCGTGGCGGATCGGC 7124

QY 1004 GAACCCGGAAGCGCTTACCGTTTCGTTTACGTGTCGAGGCGCTCGCTCGTGTG 1063
|||||
DB 7125 CCATCTCGCGCGCGCGGACCGCTTCGCGCGATCTCGCGCGCTCCCGCGCGCGCG 7184

QY 1064 GATGCGCGTGGTAAGGCAATTTACGGTATGTCCTGTCATACCTGCTCATACGCGCA 1123
|||||
DB 7185 CGACGCGGGGTTACACCTTCACCGTATCTCTCGACACACCTCCCGCGCAACGGCCA 7244

QY 1124 TGTGCGCGCAAGGCTTCTCGCTGATGTCGAAGCGCGGATCGGAGTTGGCGAAGTG 1183
|||||
DB 7245 TGTACCCGCAACGCGCTGTTGGGCTGCGGAGCTCTACGACGCGAGCTTGGCGGTG 7304

QY 1184 GATTGAGAAACCGGACCTTCCCGAAGCGGAATGTTGATCGATCAACCCGACCGGTTTC 1243
|||||
DB 7305 GGTGAAGCGCAGGTGGCTTCCCGAAGCGGATGTCGACCGCATCACCCCGCGCACCGG 7364

QY 1244 GCGGAATTCGCAAGAGCTCAACCGCGGCGAGTGGGCTGATGACGACCTGCCGCTGGT 1303
|||||
DB 7365 CCGCA---CGAGCGCGAATGTGGCGAGGGTTCGGCTCGCGATCGGTTGCGCGTCAAC 7421

QY 1304 GCGCGAGGATTTCCATGAGTGGTCTGGAAGACCAAGTTTCGGATGCGCGTCCGCCCT 1363
|||||
DB 7422 CTGCGAGCGGTTCCGGAGTGGGTGATCGAGATCATTTCCCGCGGACGCGCGCGCT 7481

QY 1364 TGAAGAACCGCGCTGACAGTGGTGGGAGCGTACGACGACTGGGAGTACGTCAGATCCG 1423
|||||
DB 7482 CGAGAAGTGGCGGTGACCTTCCACCGCGCATGTCTCATGCTACGAGGCGATGAAGATCCG 7541

QY 1424 AATGCTCAATGAGGCGATGTCATGCTCTGCTTCCAGGCAATTCCTGCTCGCTATGAGA 1483
|||||
```

```
Db 7542 CATCCTGAACGGGGGCATCGCGTGATCGCTATCCCTCGCGGCTCATGGACATCCAGCT 7601

QY 1484 TGTGGATGACGCCATTGAAGACACGCAACTCCTTGGCAATCTGAAGAACTATCTCAACAA 1543
|||||
Db 7602 CGTGCAGCGGCCATCGCCATCGCTGATCGCGGCTTCTCTGCACAGTTCGAGGTGCGA 7661

QY 1544 GGATGTCATCCGACCCCTGAAGCGGCTTTCAGGATGACGCTCGAAGGCTATCGGGACAG 1603
|||||
Db 7662 GGAGATCCTCGCCCATGTCGCGCGCTGCGCGACACAGCATCCCGACTATCTTACCCCT 7721

QY 1604 CGTCATCAGCGGTTCTTCCAAAGGCGATGTCGGACACGACGCTCCGGATGTCAGCA 1663
|||||
Db 7722 GATCGAGAGCGGCTTTCGAACCCCGAGATCGCGACACGACGCGGAGGCTCTCCCTCGA 7781

QY 1664 TGGCTGTTCCTCAAGGTTTCAGGTGTTTC 1688
|||||
Db 7782 CGGTTCGAACCGGCGAGGTTTC 7806
|||||

RESULT 13
AE005663 10955 bp DNA linear BCT 21-MAR-2001
LOCUS Escherichia coli O157:H7 EDL933 genome, contig 3 of 3, section 282
DEFINITION of 290.
ACCESSION AE005663 AE005174
VERSION AE005663.1 GI:12519330
KEYWORDS
SOURCE Escherichia coli O157:H7 EDL933.
ORGANISM Escherichia coli O157:H7 EDL933
Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae; Escherichia.
REFERENCE 1 (bases 1 to 10955)
Perna,N.T., Plunkett,G. III, Burland,V., Mau,B., Glasner,J.D.,
Rose,D.J., Mayhew,G.F., Evans,P.S., Gregor,J., Kirkpatrick,H.A.,
Posfai,G., Hackett,J., Klink,S., Boutin,A., Shao,Y., Miller,L.,
Grotbeck,E.J., Davis,N.W., Lim,A., Dimalanta,E., Potamouisis,K.,
Apodaca,J., Anantharaman,T.S., Lin,J., Yen,G., Schwartz,D.C.,
Welch,R.A. and Blattner,F.R.
Genome sequence of enterohaemorrhagic Escherichia coli O157:H7
Nature 409 (6819), 529-533 (2001)
2 (bases 1 to 10955)
Perna,N.T., Plunkett,G. III, Burland,V., Mau,B., Glasner,J.D.,
Rose,D.J., Mayhew,G.F., Evans,P.S., Gregor,J., Kirkpatrick,H.A.,
Posfai,G., Hackett,J., Klink,S., Boutin,A., Shao,Y., Miller,L.,
Grotbeck,E.J., Davis,N.W., Lim,A., Dimalanta,E., Potamouisis,K.,
Apodaca,J., Anantharaman,T.S., Lin,J., Yen,G., Schwartz,D.C.,
Welch,R.A. and Blattner,F.R.
Direct Submission
Submitted (22-OCT-2000) Laboratory of Genetics, University of
Wisconsin, 445 Henry Mall, Madison, WI 53706, USA
FEATURES
Location/Qualifiers
1. 10955
/organism="Escherichia coli O157:H7 EDL933"
/strain="EDL933"
/serotype="O157:H7"
/db_xref="taxon:155864"
/note="enterohaemorrhagic"
complement(118..1461)
/gene="gntP"
/note="Z5919"
complement(118..1461)
/gene="gntP"
/function="transport; Transport of small molecules:
Carbohydrates; organic acids, alcohols"
/note="Residues 1 to 447 of 447 are 100.00 pct identical
to residues 1 to 447 of 447 from Escherichia coli K-12
Strain MG1655: B4321"
/transl_table=11
/codon_start=1
/product="glucuronate transport system permease 3"
/protein_id="AAG59503.1"
```

```
/db_xref="GI:12519331"
/translation="MHVUNILWVFGIGLMLVLNLFKFKINSWMAHTLLVAALSVMGLAGM
DLMSLLHMKAGFGNTLGEALIVVFGAVIGKLMVDSGAHQIAVHTLIARGLRVOL
SVIIGLFLGFAMFYVAFIMLAPLIVIAAEAKIPFLKLAIPAVAATAATFSLFPPO
PGPVALNAYGADCMWYVYGVLTVPISVICAGLILPKFLGNLERPTPSFKADQPDV
MNNLPSGVSLVPLIPALIMISTIANIWLKOTPAWEVNFVIGSSPIAFIAKVA
FVLGTARGHDMQWMMNFAFSAKSIAMVILIIGAGGVKOTIIDTIGDGTIGMLMSH
GNISPYIMAWLITVIRLATCGVVSAMTAAGIISAILDPATGOLGVGNPALLVLAT
AAGSNTLTHINDASFMLFKGYFDLSVKDTLKTWGLLELVNSVVGIIIVLIISMVA"
:
gene
CDS
/compound="enzyme; Degradation of small molecules: Carbon
compounds"
/function="enzyme; Degradation of small molecules: Carbon
compounds"
/note="Residues 1 to 393 of 394 are 99.23 pct identical to
residues 1 to 393 of 394 from Escherichia coli K-12 Strain
MG1655: B4322"
/codon_start=1
/transl_table=11
/product="mannonate hydrolase"
/protein_id="AAG59504.1"
/db_xref="GI:12519332"
/translation="MEQTRWYGNPDVSLADVROAGATGVVTHLHHPNCEVSVSE
ILKRAIVEDAGLVSVSVFPIHEDIKHTGNYEQMTANYOQTLNLAQCGIRTVCY
NFMVLDWTRDLEYVLPDGSKALRFQIEFAAFEMHILKRPGEADYTEEIAQAAE
RFATMSDDEKARLNRIIAGLPGAEGYTLQDFRKHLELYKDIDKAKLRNFVFLKA
IIPVAEEVGMVAPHPDPRPRIILGPRIVSTIEDMOMVDTVNSMANGFSMCTGSYG
VRADYDLVMIKQCPRIYFTHLSTMEEDNPKTFEAAHLNGDVMYVYKAIVEEE
HRRKAEKEDLIPRPDGHOMLDDLKKTNPGISAGRLAGLAEVRGVELATORAFF
SH"
gene
CDS
/compound="enzyme; Degradation of small molecules: Carbon
compounds"
/function="enzyme; Degradation of small molecules: Carbon
compounds"
/note="Residues 1 to 486 of 486 are 99.58 pct identical to
residues 1 to 486 of 486 from Escherichia coli K-12 Strain
MG1655: B4323"
/codon_start=1
/transl_table=11
/product="mannonate oxidoreductase"
/protein_id="AAG59505.1"
/db_xref="GI:12519333"
/translation="MTTIVDSNLVPARSDHRSLEIRVILHGCAPFRAHOALYTHH
LLESTDSDWGCIVENLMPGNDRVLIENLKKOQLLYTVAEKGAESTELKIGSMKEALH
PEIDCEGILNAPARQPAIVSLVTEKGYCADAASGOLDLNNPLIKHLENPAKPS
AIGYVEALRREKGRKAFVTMCDNVRENGHVAKVAVLGLAQARDPQLAAANTEENY
TFPQTMVDRIYPAATPETLQIADQLGVYDPCAICEPFRQWIEDNFVNGRPDWMKV
GAQFVADVPEEMKRLMNGSHSLFAYLGYLGVEYETIADTMTPDYKKAFAALMVOE
QAPTLSPGEGDLNAYATLLTERFSPNSLRHTWQIAMDGSKLPORLLDPVRLHON
GGSWRHLALVAGHMRVYTOGVDEOGNALDVYDPMLEAFQKINAQYQAGDRVKALLGLS
GIFADDLPQNAADVGAVTAATAQQLCERGARECVAAI"
4742..5515
/gene="uxuB"
/note="25922"
4742..5515
/gene="uxuB"
/function="regulator; Degradation of small molecules:
Carbon compounds"
/note="Residues 1 to 257 of 257 are 98.83 pct identical to
residues 1 to 257 of 257 from Escherichia coli K-12 Strain
MG1655: B4324"
/codon_start=1
/transl_table=11
/product="regulator for uxu operon"
/protein_id="AAG59506.1"
/db_xref="GI:12519334"
/translation="WKSATSQRPYQYQXAIRDLIIKTPNPGERLPPPERIAEMLD
VTRTVREALMLEIKGLVEVRGAGIYVLDSGSHNTDSPDANVCNDAGPFELLQAR
```

```
OLLESNTAEFAALQATREDIVKMQALQLEERELASSAPGSESDGMQFHIAIEATH
NSMVLVFRSQWQRRNPMWIKLHSDLSLYRKEMGLDHIKQILAALIKKARAARKL
AMWOHLENKWLFEFSNVDDIYFDGYLDSWPLDKVDA"
complement(5656..6486)
/gene="yjiC"
/note="25923"
complement(5636..6486)
/gene="yjiC"
/function="orf; Unknown function"
/note="Residues 1 to 276 of 276 are 93.47 pct identical to
residues 1 to 276 of 276 from Escherichia coli K-12 Strain
MG1655: B4325"
/codon_start=1
/transl_table=11
/product="orf, hypothetical protein"
/protein_id="AAG59507.1"
/db_xref="GI:12519335"
/translation="MSMPLSNVLSQIITDNHFLHHPKVESLPRKYERARLDTENIY
LLPLARGNNHNDGKSVVEIRKLDISKESWPNFYVTEACRESDGITTTGRMLYRNKI
TSALDEIYGICRKAHAATELAEGLRLNLFKSPDPVEDYTVHEITLPGCNVPGYA
GTTIGYSTLPASOKAKRWTEQPRIDIVDOIITVSGVANSNGFALALLNANIELGN
DPIIGIEAYPGTAETIAKMGYKVIPGDEDAPLKRWTLQPSLLPELFLKNGEWNVYIG
"
6820..6963
/gene="25924"
6820..6963
/gene="25924"
/function="orf; Unknown function"
/note="No significant matches"
/codon_start=1
/transl_table=11
/protein_id="AAG59508.1"
/db_xref="GI:12519336"
/translation="MADYAEINNPPPELSSGDKYFHLRNYSEYSEYTSGFFLSLMIF
IKS"
7150..7551
/gene="yjiB"
/note="25925"
7150..7551
/gene="yjiB"
/function="orf; Unknown function"
/note="Residues 1 to 133 of 133 are 98.49 pct identical to
residues 1 to 133 of 133 from Escherichia coli K-12 Strain
MG1655: B4326"
/codon_start=1
/transl_table=11
/protein_id="AAG59509.1"
/db_xref="GI:12519337"
/translation="MRKMMROSQAVALPEISGNKTSLLRKSVCSDLLTLFNSPHSTLP
SLLYSGMPEQWVHNPSDKHLQSWYCRQLRSALLFHEPRIAALQVNLKRAYCHTLAISL
ETIMLYHDDPELTFDLVMDNGGWSATLENVS"
complement(7544..8272)
/gene="yjiE"
/note="25926"
complement(7544..8272)
/gene="yjiE"
/function="putative regulator; Not classified"
/note="Residues 1 to 242 of 242 are 98.34 pct identical to
residues 62 to 303 of 303 from Escherichia coli K-12
Strain MG1655: B4327"
/codon_start=1
/transl_table=11
/product="putative transcriptional regulator LYSR-type"
/protein_id="AAG59510.1"
/db_xref="GI:12519338"
/translation="MTPLQSEQCKFIHFSQIRHLLQQLAESNLAELRGSSDYAQRKIKI
AAHSLSLGLPLSIISQMPPLFTWAIEAIDVADVKLRGOSDCIFSFHDEDLLEAP
FDHRLRFESQLFPVCASDEHGFALDLQPHFPLNYSRNSYMGRLINRTRHSELS
FSTFPVSSMSELLKQVALDGGCIAWLPEYALQQEIRSQLVVLNRDELVIPTQAYAYR
MNRMPNVAERFWRELDELVLS"
complement(8520..9692)
/gene="fadA"
/note="25927"
```

CDS	complement(8520..9692) /gene="iada"				
Query Match	5.8%;	Score 236.8;	DB 1;	Length 10955;	
Best Local Similarity	49.7%;	Pred. No. 1.3e-37;			
Matches 604;	Conservative 0;	Mismatches 612;	Indels 0;	Gaps 0;	
Qy	616	GGATCGTGCAATTCGGGTAGTAGTAATTTTTTCGAGCCCATGAGCGGCTTCTACGTGCGAGC	675		
Db	3137	GCATTTGCAATTCGGTTGCGGGCGCTTTACCGTGGCACCAGGCGCTGTACACCATTC	3196		
Qy	676	AGATTCTTGAAACCGCTCCGGACTGGCGGATTTGTTGTGGCCTGACGGCAGTGACC	735		
Db	3197	ATCTGCTGGAAACGACCGACAGCGACTGGGGTATCTGGAAATTTACCTGATGCCAGGCA	3256		
Qy	736	GTTTCAAGAAAAAGCGAGGAATTC AAGGCCCGAGCACTGCCTGTATTTCCCTGACCGAGA	795		
Db	3257	ACGACCGCTGCTGATCGTGAACAACTGAAAAACAGCAACTGCTGTACACGTTGGCGAAA	3316		
Qy	796	CGGCTCGGTCGCGAAGACACGTCGGCGTCATGGCGCGCTGCGTGACTATCTGCTTG	855		
Db	3317	AAGCGCGGAAGCACCGAGCTAAAAATTTATCGGTTGGAAGAAAGCGCTGCACCCCG	3376		
Qy	856	CCCGCGCGATCCGGAAGCGCTGCTGAAGCATCTTGTGATCCGGCGCATCCGCATCGTTT	915		
Db	3377	AAATCGACGCTCGGAAGGTATTTCTCAACGCGATGGCGCTCCGCACAAACGGCGATTGTCT	3436		
Qy	916	CCATGACGATCAGGGAAGGCGGTACAACATCAACGAGACACCGGTGCTTCGATCTGG	975		
Db	3437	CCCTGACGGTTCAGGGAAGAGGCTACTGCGCTGATCGGCAACGCGTCAACTGGATCTCA	3496		
Qy	976	AGATCGGCGAGTAAAGGCGCGACCTCAAGAACCCGGAAGCGCTCTACGTTTTCGGTT	1035		
Db	3497	ATAACCCGCTGATCAAGCAGCATCTGAAAAACCCGACTGCGCGCGAAGTCCGCGATGGTT	3556		
Qy	1036	ACGTGTCGAGGCCCTCGCTCGTGTGGATGCCGTTGTAAGGCAATTTACGGTCAATGT	1095		
Db	3557	ACATCGTCGAAGCCTTCGCTGCTGCTGCTGAAAAAGGGCTGAAAGCATTTTACGGTATGT	3616		
Qy	1096	CCTGTGATACCTTCGCTCATTAAGGCACATCTGCCCGCAAGGCTTCTCCGCGCTATCGGA	1155		
Db	3617	CCTCGGATACTGCTGTAAGACGGTCATGTGGCGAAGGTCTCGGCTACTGGGCTGGCTC	3676		
Qy	1156	AGGCGCGCATCCGGAGTTTGGCAAGTGGATTGAGGAAAAACGGACCTTCCCGAAGCGAA	1215		
Db	3677	AGCGCGTGAACCGCAGCTAGCGGCATGGATTGAAGAGAACGTCACCTTCCGTGACCA	3736		
Qy	1216	TGTTTGATCGCATCACCCGACCGTTTCGCGGGAATTCGCCAAGAGCTCAACGCGGCCA	1275		
Db	3737	TGTTTGACCGCATGTTTTCGCGGCGGACGCCAGAACTTTACAGGAATTTGCTGACACAGC	3796		
Qy	1276	GTGGGCTGGATGACGACCTCGCGCTGTGTGGCCGAGGATTTTCCATCAGTGGGCTCTGGAAG	1335		
Db	3797	TGGTGTTTACGACCCGTGCGCCCATTTGCCTTGGNACCGTTCCGTCAGTGGGTTATTCAG	3856		
Qy	1336	ACCAAGTTTGGGATGCGCGTTCGCGCGCTTGAAAAAGCCGCGTGCAGATGGTCGGGACG	1395		
Db	3857	ATAACTTCGTTAATGGTCTGCGGACTGGGATAAAGTGGGCGCACAGTTTGGTTGACAGC	3916		
Qy	1396	TGACGGACTGGGAGTACGTCAAGATCCGATGCTCATGCAAGGCGATGTCATGCTCTGCT	1455		
Db	3917	TTTGTGCGGCTCGAAATGATGAACCTGCGTATGCTGTAACGGGACGCACTTTTCTGGCGT	3976		
Qy	1456	TCCCAGGCAATTCCTGCTCGGCTATGAGAATGTGGATGACGCCATTTGAAGACAGCGCACTC	1515		
Db	3977	ACCTCGGTTACCTCGCGCGCTATGAACATATTGCCGACACCATGACTTAACCCGATTATC	4036		
Qy	1516	TTGGCAATCTGAAGAATATCTCAACAAGATGTCATCCGACCCCTGAAAGCGCCTTCAG	1575		
Db	4037	GCAAAGCGGCTTTTGGCCCTGATGATGCAGGAACAAAGCGCAACGCTGTGATCGCGGAAG	4096		
Qy	1576	GCATGACGCTCGAAGGCTATTCGGGACAGCGTCAATCAGCCGTTTTCTCCAAACAAAGCGCAT	1635		

TITLE JOURNAL	Direct Submission	
	Submitted (26-JUN-2000) Ken Kurokawa, Osaka University, Genome Information Research Center; 3-1, Yamadaoka, Suita, Osaka 565-0871, Japan (E-mail: ken@gen-info.osaka-u.ac.jp, URL: http://www.gen-info.osaka-u.ac.jp/, Tel: 81-6-6879-8365, Fax: 81-6-6879-2047)	
COMMENT FEATURES	genome project.	
	Location/Qualifiers	
source	1..168764	
	/organism="Escherichia coli O157:H7"	
gene	/strain="O157:H7"	
	/sub_strain="RMD 0509952"	
CDS	/db_xref="taxon:83334"	
	79..243	
gene	/gene="ECS5227"	
	79..243	
CDS	/gene="ECS5227"	
	/note="identical to B4250_ECOLI gi11790699 (Conserved in E.coli K-12)"	
gene	/codon_start=1	
	/evidence=not_experimental	
CDS	/transl_table=11	
	/product="hypothetical protein"	
gene	/protein_id="BAB38650.1"	
	/db_xref="GI:13364705"	
CDS	/translation="MLSVLLKLLILVLTPTASTRLLAVALGYLAVYSMMNSVYRKLFR LSIFVMIYR"	
	315..569	
gene	/gene="ECS5228"	
	315..569	
CDS	/gene="ECS5228"	
	/note="similar to YJGJ_ECOLI gi11790700 percent identity 92 in 84 aa (Conserved in E.coli K-12)"	
gene	/codon_start=1	
	/evidence=not_experimental	
CDS	/transl_table=11	
	/product="hypothetical protein"	
gene	/protein_id="BAB38651.1"	
	/db_xref="GI:13364706"	
CDS	/translation="MVLEGIHSDHPQARDIAVQYHAAETAIDYIARLHPQSQCVT DFMSTVMSGLSAKARECHSLQLCATAALAGEAIKTIKE"	
	705..1166	
gene	/gene="ECS5229"	
	705..1166	
CDS	/gene="ECS5229"	
	/note="similar to YJGK_ECOLI gi11790701 percent identity 98 in 153 aa (Conserved in E.coli K-12)"	
gene	/codon_start=1	
	/evidence=not_experimental	
CDS	/transl_table=11	
	/product="hypothetical protein"	
gene	/protein_id="BAB38652.1"	
	/db_xref="GI:13364707"	
CDS	/translation="MEKMIIGNIINLPWLPOELRQALEHKAHVTAETPKGKHIDEG NRFLYLSMEDTEPYEARRYHARYLDIQVLKGQEGMTFTQPAGTPDPTDWLADKD IAFILPEGVDEKTVILNEGDFVVFYPGEVHKPLCAVGAQPAQVRKAVVKMLMA"	
	1289..3061	
gene	/gene="ECS5230"	
	1289..3061	
CDS	/gene="ECS5230"	
	/note="similar to part of (13-616 in 616 aa) YJGL_ECOLI gi11790702 percent identity 82 in 590 aa"	
gene	/codon_start=1	
	/evidence=not_experimental	
CDS	/transl_table=11	
	/product="hypothetical protein"	
gene	/protein_id="BAB38653.1"	
	/db_xref="GI:13364708"	
CDS	/translation="MSKISDLNYSQHITLADNFKOKSEVLNTRVGMNNTFARNAGGOD NTRNLNPKTLEFLVKIFTIGYVDFSKRSNAGRNMAHLESSSYIKNNDGSPIMKF VMNPEGERADSPKVIIEISLITTTWGTQGHATLIIFPOPDGSTNRYERKSPERKDE SSLHLITKNVLACVQRANKREIARLNHOKLNQNLNLOKLNLOKLNLOKLN IQELNNSQELNNSQELNNSQDLKNSQVSKGSDVDTITDILLEKSLNNALLAI	

RNEHLLMPHYVCSESI SYLLGENGILEEIDKLYELNDHGIDNDKEGNEINDIMINLS
HILIESDADKVNLTPIVHSMLTFFLELPYNNDVKILEWCFNKSMQYFDDSAKLEHAC
SVINIFNRDQSKVAETLEFNLDKEPKNSPELOELIMKLVVYVNDNLSNREKTV
LIQRIFFNVSLEFNKVPVSLVNDIEFNDFEMKNTMINNYFPLLSYEDEKLYFDK
LGYNFNKNSEETMKNOQPDVIEEKLNNELKLFRRMMQITLKSEVVNSPPIQORLN
TLNPPENLRITAEKFGWKKTTITA
/complement(3118..4122)
/gene="ECS5231"
/complement(3118..4122)
/gene="ECS5231"
/note="similar to ARGF_ECOLI gi11786469 percent identity 87 in 334 aa, similar to ARGI_ECOLI gi11790703 percent identity 99 in 334 aa (Conserved in E.coli K-12)"
/codon_start=1
/evidence=not_experimental
/transl_table=11
/product="ornithine carbamoyltransferase 1"
/protein_id="BAB38654.1"
/db_xref="GI:13364709"
/translation="MSGFYHKHFLKLLDTPAELNSLLQLAALKADKKSKEEAKLT
GKNALIFEKDSRTTRGSEFVAAYDOGARVYTLGPGSGQIGHKESIKDTRVLGRMYD
GIOYRGYQETIVETLAVAGVPVWNGLTNEPHTOLLADLLTMQEHLPGRPFNEMTLV
YAGDARNMGNMSEAAALTCLDLRLVAPQACWPEAALVTECRALAOQNGNITLTED
VAKSGVEGADFIYDVVWYMSGEAKKWAERIALDRDYQVNSKMMOLTGNPEVFLHCLP
AFHDDQITLGRKKMAEEFGLHGMETDEVFESAASIVFDQAEENRMTIKAVWVATLSK
"
4284..4700
/gene="ECS5232"
4284..4700
/gene="ECS5232"
/note="similar to YJGD_ECOLI gi11790704 percent identity 100 in 138 aa (Conserved in E.coli K-12)"
/codon_start=1
/evidence=not_experimental
/transl_table=11
/product="hypothetical protein"
/protein_id="BAB38655.1"
/db_xref="GI:13364710"
/translation="MANPEOLEQREERTLRIEELLEDGSDPDALTYTIEHLSADDLE
TLKAAVAEPKIGYEVTDPELEVEDGDIVTCCLISECALNADLIDAQVEQLMTLAE
KFDVEYDGGWGYFEDPNGEDGDDFVDEDDDGVRH"
complement(4746..5039)
/gene="ECS5233"
complement(4746..5039)
/gene="ECS5233"
/note="similar to B4256_ECOLI gi11790705 percent identity 93 in 97 aa (Conserved in E.coli K-12)"
/codon_start=1
/evidence=not_experimental
/transl_table=11
/product="hypothetical protein"
/protein_id="BAB38656.1"
/db_xref="GI:13364711"
/translation="MVGGGTAPLAGSESDICELOKMYFLPAIRGKGLAKKLALKAME
EARDMGKRYLETTFATLKEAIGLYEHLGHQIDYALGCTGHVDCEVRMLREL"
5442..6638
/gene="ECS5234"
5442..6638
/gene="ECS5234"
/note="similar to part (5-393 in 393 aa) of YJGN_ECOLI gi12367367 percent identity 96 in 389 aa (Conserved in E.coli K-12)"
/codon_start=1
/evidence=not_experimental
/transl_table=11
/product="hypothetical protein"
/protein_id="BAB38657.1"
/db_xref="GI:13364712"
/translation="MAOVINEMDVPVSHSFVHGTERGYFLICVVNLLTITIGIYLP
WALMKCKRYLYANMEVNGQBFYGITGNNVFESCLVFFVFFAILMTVSADMPILGCV
LTLSELLVLLIFMAAKGRYQALMTSLNGRFRFSNCKMKGWVWVTFPLILMAGMGTV
FFTSTKMLHANSSSVISVVVMAIVGIVSIGIFNGTLYSLVMSFLWSNTSFGTHREK
VKLPTAYCIKYAILAFIALFLAVAGYIIFDQILNAYDSSVYANDDIENTLQOFMEQO


```
/db_xref="GI:15141287"
/db_xref="SPTREMBL:Q92TU6"
/translation="MMDGASNAAPRIIRGHPALPDLSDGRVALITGSSGGGLTMAHALC
EAGSVILNGDEARLAGEALEEHEGTVGTSAFDVTKSAETGRAVADILAEGRGID
ILVNNAGIORTPLHEPPEDAFRVLETNLTSALFLYQAVVOGMIERREGTINICSV
QSELARPSATPYAASKGGLKMLTKMALDWRGIRVNGLAGPGYFKTELSNALSUDEK
FSTWLSORPLGKWDGTGELAAAVFLAASFSFVGHILYVDGITSCL"
gene
CDS
complement(1063..1361)
/gene="Smb20693"
complement(<1063..1361)
/gene="Smb20693"
/function="MISCELLANEOUS; Hypothetical/Partial homology"
/notes="Product confidence : hypothetical
Gene name confidence : hypothetical"
/codon_start=1
/evidence=not_experimental
/transl_table=11
/product="HYPOTHETICAL PROTEIN"
/protein_id="CAC49801.1"
/db_xref="GI:15141288"
/db_xref="SPTREMBL:Q92TU5"
/translation="MPDYAASKGVGVGLAKAFATNGPGTFRSMPLPPGYIATDSPEA
SRNDPSRAESILARIIPRAAGAGPRTLPAARRSFWRHPITSRASSPWTAAAGWA"
gene
CDS
complement(1508..1684)
/gene="Smb20694"
complement(1508..1684)
/gene="Smb20694"
/function="MISCELLANEOUS; Hypothetical/Partial homology"
/notes="Product confidence : hypothetical
Gene name confidence : hypothetical"
/codon_start=1
/evidence=not_experimental
/transl_table=11
/product="HYPOTHETICAL PROTEIN"
/protein_id="CAC49801.1"
/db_xref="GI:15141289"
/db_xref="SPTREMBL:Q92TU4"
/translation="MGADGDPMAALDARMVRVIERLRVADMSAVPDINAGNTNAPAT
MLGNRCAGFILGTA"
gene
CDS
complement(1765..2271)
/gene="Smb20695"
complement(1765..2271)
/gene="Smb20695"
/function="MISCELLANEOUS; Unknown"
/notes="Product confidence : hypothetical
Gene name confidence : hypothetical"
/codon_start=1
/evidence=not_experimental
/transl_table=11
/product="HYPOTHETICAL PROTEIN"
/protein_id="CAC49802.1"
/db_xref="GI:15141290"
/db_xref="SPTREMBL:Q92TU3"
/translation="MTLLPVALGGTELVAWRLDQAVHAPTWDSGEGAYRVGGRWNSKG
VRAYVCSLPATVAILLEVAVHKGFRALDVAHTWTAAADAGDVYVYVDPKSPVNPWA
RCPIGAGQQAQVGGDILRRHREVAIPSAVSPHSWNLVFLGGAAAPALAKFOEPFALD
TRLHPPGT"
gene
CDS
complement(2268..2753)
/gene="Smb20696"
complement(2268..2753)
/gene="Smb20696"
/function="MISCELLANEOUS; Unknown"
/notes="Product confidence : hypothetical
Gene name confidence : hypothetical"
/codon_start=1
/evidence=not_experimental
/transl_table=11
/product="HYPOTHETICAL PROTEIN"
/protein_id="CAC49803.1"
/db_xref="GI:15141291"
/db_xref="SPTREMBL:Q92TU2"
/translation="MALAKAPTPAAGGGELOKIEITLLGGSRVLAHLNLDAHELL
LRGLPASADHLVGNLVFIKGTESLEKAVGMSLRTWQRKTEPNKPLSQEQSGRAWKF
A"
gene
CDS
complement(2855..4549)
/gene="Smb20697"
complement(2855..4549)
/gene="Smb20697"
/function="MACROMOLECULE METABOLISM; Macromolecule
degradation; degradation of proteins, peptides,
glycopeptides"
/notes="Product confidence : putative
Gene name confidence : hypothetical"
/codon_start=1
/evidence=not_experimental
/transl_table=11
/product="putative peptidase, similar to arginine
utilization protein RocB"
/protein_id="CAC49804.1"
/db_xref="GI:15141292"
/db_xref="SPTREMBL:Q92TU1"
/translation="MTNAQEETVSRAQSIDSERVQFALRMTSPSETGTGGEASFA
DRHLGLLGLPYFREHPDGLLASHGEPLTRNNVALVRGTGKRTLVMAGHFDVSTFD
NYHELKALACDSLALKDALTESLSARTGRSEQERALQDLASGDFLPGRGLLDMKSL
AVAIACLEQFAADTDROGNMLVATPDEERESGRMSLRDALPGLVRDEIDETAGIN
LDVTSQDGSSEGRAVIAGTICKLDRKLDLPGYEVTPFERFIANLWYASQAAAGLLARE
GNASLADRDNDISPPPICLEAKDLRDYEVTPFERFIANLWYASQAAAGLLARE
EEVLTGANEAIETFAQSAEYKGLVGRSAGLPAKPRLLSFOELRAAAARVFGHGFDA
FYAEKREFAQSDNPLVATROLTEMLVGRSLGPAIVTIGFSGHLYPPSHRLAECND
RSLHOATEKARAGLNDPARSLLIWKPHFYGISDMSFLGAASGVSVSDNTPISRLVD
RSENALRFTYVNLGPGWREFHQKPERVHPYAFRVLPVELVSEIARTFLCDDHRD"
gene
CDS
complement(4701..6353)
/gene="Smb20698"
complement(4701..6353)
/gene="Smb20698"
/function="CELL PROCESSES; Transport of small molecules"
/notes="Product confidence : putative
Gene name confidence : hypothetical"
/codon_start=1
/evidence=not_experimental
/transl_table=11
/product="putative transport protein, similar to E. coli
multidrug resistance protein EmrB"
/protein_id="CAC49805.1"
/db_xref="GI:15141293"
/db_xref="SPTREMBL:Q92TU0"
/translation="MTIATATPSRIDSMAGLLVAGIVLATLEATASTVLSLRGDI
IGTYATPDEFALWDVGYTTFKILIGMAASWLMRDPRLNVVSGSTLAMGACGIAAM
TARLDLLVALRMIOGSPGGTLLVGGQAILFLTFPHSRQPLLOAFAMGSVAPATAP
ALQGLLLDSQSWTWIFFSIIPALAAAGLLLLADGPVATRAQRPPDWGFSLSAAL
FCFTYVFSQGSQWDFEERILWLAVIGTATLTLAFIQQOVLKANGQGLFDTLTFEDF
CFATVSVGAALFGSAFLIPSAFVSVLAFTPDAGLLPSGALFLGALLIAEFLM
QLRRVPPVATVPFILIMAAWMLSGSTSESGADGMAAILIRGALGFLGLISITLI
AFSNLSRNLASGILGTLTGROLGLIGVSAIQTLIENHVSHNVLAVLGANVTAGAPAV
ADRLTTAALLTAKGMDAAAASRGASLLRGAVAGOSTVIAFDATFNAIGLLFVIAAP
VLVGKIGILARYAKARAEKGRAGVAAKPSCHHTLPKRPAS"
gene
CDS
complement(6350..7447)
/gene="Smb20699"
complement(6350..7447)
/gene="Smb20699"
/function="CELL PROCESSES; Transport of large molecules;
protein, peptide secretion"
/notes="Product confidence : putative
Gene name confidence : hypothetical"
/codon_start=1
/evidence=not_experimental
/transl_table=11
/product="putative protein secretion protein, HlyD family,
similar to E. coli EmrK"
/protein_id="CAC49806.1"
/db_xref="GI:15141294"
/db_xref="SPTREMBL:Q92TU9"
/translation="MESLEKVAITGARGSGNCVAVARNRKKMGICLALCLGAVVVVV
AGWTWARESAASTNAYVRGDTVTSIAPKAVGYTAVEVEQAVRAGDVLFRIDRD
YRAKLAQAVANVFAAEARLTNVDAEMALQHALIROAQRRSVVAELNLAARAYRRR
```

ELIRSETISQAHVDESDAKSRAEANVLAASATVEAQOORIAVLAAQREAAVAQAQ
EAARDLAGIDLESTVVRAPVGVGNQVRVGRVAPGASLLDIVPLDNWIVANEKE

Query Match 5.78; Score 236.2; DB 1; Length 189333;

Best Local Similarity 50.78; Pred. No. 1.3e-37;

Matches 655; Conservative 0; Mismatches 618; Indels 18; Gaps 3;

Qy	539	GATTACCGCGGAAACCCCTTAAGTCTCTCTCTGCAATGTCCAGGCTCCCGCCCTATGACAT	598
Db	44081	GCITTCGAGAAGACGATGACCCCTGCGCGCACCGTCAAAAGGCCCATTTACGACCT	44140
Qy	599	CGACGGGATCAAGCCTCGGATTCGTCATTTTCGGTGTAGGTAACTTTTCGAGCCCCATGA	658
Db	44141	CGGACAGATCACCGTCGCGCATCGCCACCTCGGCATCGGTGCTTTTCATCGCGCGCACCA	44200
Qy	659	GGCGTTTACGTCCAGCAGATTTTGAACAGCTCCCGGACTGGCGATTGTTGGTGTGG	718
Db	44201	GGCGTCTACACCATGGGTGCTCTCCGAGGATCGTCTGGGGCATCTCGCGCGTCTC	44260
Qy	719	CCTGACGGGAGTGACCGTTTCAAGAAAAAGCCGAGGAATTCAGGCCCCAGGACTGCCT	778
Db	44261	GCTCGGACGCCGAG-----ACGCGGACGCGCTTCATCCGACGAGCGGCT	44308
Qy	779	GTATTCCCTGACGAGACGGTCCGTCGCGCAAGACACGGTGGCGTCAATGGCGCGCT	838
Db	44309	CTACACCCCTCGCGTCCAGGACGGCGAAGGCTC---CGAGCTCAGCGTGGTTCGGCAGTGT	44365
Qy	839	CGGTGACTATCTGCTTCCCGCGCGGATCCGGAAGCGGTGCTGAAGCATCTTGTGATCC	898
Db	44366	CGTCGAGTCTTTCGCGCGGAGACCGCGGAAGCGGCTCTTCGCGCGATGGCGGATCC	44425
Qy	899	GGCCATCCGATCTGTTTCCATGACGATCAGGAGGGGCTTACAACATCAACGAGACGAC	958
Db	44426	CGGCACACGGATCGTGTGACGATCACCAGAAAGGCTATTGCCATTAACCCGCGAAC	44485
Qy	959	CGGTGCGTTCATCTGGAGAAATCGGCGAGTAAAGCCGACGCTCAAGAACCCCGAAAAGCC	1018
Db	44486	CGGCACGCTCGAAGAGGCCACCGGACATCGTCCAGCACTCGCAAAACCCCGCGCGCC	44545
Qy	1019	GTCTACCGTTTTCGTTACGTGTCGAGGCCCTCGCTGCTGGTGGGATCCGGTGTAA	1078
Db	44546	GCATCTCGGATCGGGTTCATCTGCGAGGGGATTTTCGCGCGCGCTCAGCGCGGTATCGC	44605
Qy	1079	GGCATTTACGGTCTATGCTGTGATAACCTGCTGATCAACGGCAATGTCGCCCCCAAGGC	1138
Db	44606	ACCCCTCACGCTCTCTCTCGATATCTTCCGGGGAATGGCCATGCTCTGAAGCGTAT	44665
Qy	1139	CTTCCCTCGGCTATCGGAAGCGCGGATTCGGAGATTGGCGAAGTGGATTGAGGAAAACGC	1198
Db	44666	CGTCACGCGATTCGCGGAAGCGCGATCCCGCCCTCGCGCGCTCGTCCGCAACGTCGC	44725
Qy	1199	GACCTTCCCGAAGATGTTGATCGCATCACCCGACCGTTTCGGCGGAAATCGCCAA	1258
Db	44726	GTCCGCTTCGACC---ATGGTGGATCGATCTGTTCCGGCCACCAACCCAGCGACCGGAG	44782
Qy	1259	GAAGCTCAACCGCGCGGCTGGGCTGATGACGACCTGCGGTGGTGGCGGAGGATTTCCA	1318
Db	44783	CGCGGTCGCTCGCAATGGGGCTCAAGATGCGTGGCCGATCATGACCAACCCCTTCG	44842
Qy	1319	TCAGTGGGTGCTGAAGACCAAGTTTGGGATGGCCGCTCCCGCGCTTGAAAAGCCGCGT	1378
Db	44843	GCAATGGGTGATGAGGAGGATTTCCCGCTCGCGCGCCCGGCTTGGGAAAAGCGCGGAGC	44902
Qy	1379	GCAGATGGTGGGACGTCGAGGACTGCTCAAGATCGAATGCTCAATGCAGG	1438
Db	44903	GCTCTTCGTCGAACAGCTCTCCGCTTCGAGTTCATGAAGCTCCGGCTTCTGAACGGGAG	44962
Qy	1439	GCATGTATGCTGCTTCCAGGCAATTCGTGGTCGGGTATGAGAAATGTGATGACGCCAT	1498
Db	44963	CCATTCCACGCTTGCCTATCTCGGCTATCTCGCGGTCGCCGAGACCGTGGCGGACGCCAT	45022
Qy	1499	TGAAGACAGCGAATCTTGGCAATCTGAAGAACTATCTCAACAGGATGCTATCCCGAC	1558

Search completed: May 18, 2003, 23:45:56

Job time : 11490 secs

Db	45023	GGCAGTTGCGGGCATGGAAGCGTTGTCGAAGGGTTGATCGGCCACGAGGTGAGCCCCAC	45082
Qy	1559	CCTGAAGGCGCCTTCAGGATGACGCTCGAAGGCTATCGGACAGGCTCATCAGCCGTTT	1618
Db	45083	CCTTCGGAACCTTCGCGGCTTCGACCTTCGGCTATCGGGCGGAACCTCTTCAGCGCTT	45142
Qy	1619	CTCCAAAGGCGATGTCGACACGACGCTCCGATTTGCTAGCGATGGCTGTTTCCAAGGT	1678
Db	45143	TCGCAATCCTGCGCTTCGCCACCGCACCTGGCAGATCGCCATGACGGCTCGCAGAAGCT	45202
Qy	1679	TCAGGTGTTCTGGACGAAACCGTGGTGGGGCGATCGAAGACAAGGGGACCTGTACG	1738
Db	45203	GCAGCAGCGCTCTCGGCTCGATCCGACCGCTTCAGAGGAGGGCGCGCTATACCCG	45262
Qy	1739	TATAGCTTCGGAATTCATCTCTCGAAATGCTGCGTGGTTCGCGACGAGAAAGCGCG	1798
Db	45263	GCTGGCGCTCGGGTTTCGCGCTTGGATGCGCTATGCGCGGCTTGACGAGGCGGCGG	45322
Qy	1799	GACGTATGATCGTCCGAGCCGACTTATGGC	1829
Db	45323	TCCCATCGAGCTGCGCGATCCGCATGCTGCC	45353

Result No.	Score	Query Match	Length	ID	Description
1	4115	100.0	4115	21 AAA37430	Gluconobacter oxyd
2	955.6	23.2	1458	20 AAZ35672	Gluconobacter subo
3	247.4	6.0	9810	20 AAZ32025	Human METHI relate
4	247.4	6.0	9810	22 AAC30082	AF018073 cdNA clon
5	213.6	5.2	1509	22 AAH65091	C glutamicum codin
6	213.6	5.2	1632	22 AAF71527	Corynebacterium gl
c 7	213.6	5.2	34980	22 AAH64966	c glutamicum codin
8	118.4	2.9	3390	23 AAS93000	DNA encoding novel
c 9	118.4	2.9	3390	23 AAS93884	DNA encoding novel

XX Gluconobacter oxydans derived D-sorbitol dehydrogenase obtainable by
PT culturing its gene-transformed host cells, useful for producing
PT L-sorboside or 2-keto-L-gluconic acid as an intermediate in L-ascorbic
PT acid production
XX
PS Claim 9; Page 60-63; 72pp; Japanese.
XX
CC The invention relates to Gluconobacter oxydans D-sorbitol dehydrogenase
CC (SLDH; AAB23172) and to the gene encoding it (AAA97430). SLDH has a
CC molecular weight of about 54 kDa and catalyses the conversion of
CC D-sorbitol into L-sorboside using NADP+ as a co-enzyme. SLDH specifically
CC catalyses the oxidation of sorbitol, mannitol and arabitol, but does not
CC act on xylitol, ribitol, inositol and glycerol. The invention also
CC encompasses expression vectors and host cells comprising the
CC Gluconobacter oxydans SLDH gene, and the recombinant production of SLDH.
CC The invention further relates to a method for preparing L-sorboside by
CC contacting the recombinant SLDH with D-sorbitol; a process for producing
CC 2-keto-L-gluconic acid by contacting recombinantly produced sorbose
CC dehydrogenase and/or sorbosone dehydrogenase with L-sorboside; and a
CC process for preparing L-ascorbic acid or its alkaline earth metals salts
CC by conversion from 2-keto-L- gluconic acid. SLDH is useful for producing
CC L-sorboside or 2-keto-L- gluconic acid as an intermediate in L-ascorbic
CC acid production. The present sequence represents the Gluconobacter
CC oxydans SLDH gene.
XX
SQ Sequence 4115 BP; 833 A; 1133 C; 1270 G; 879 T; 0 other;

Query Match 100.08; Score 4115; DB 21; Length 4115;
Best Local Similarity 100.08; Pred. No. 0;
Matches 4115; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 AAGCTTGATGCTCCAGGTGACCTAGAGATCCGGTTTGGCAGCGCTCCCTAGATT 60
DB 1 AAGCTTGATGCTCCAGGTGACCTAGAGATCCGGTTTGGCAGCGCTCCCTAGATT 60
QY 61 GATCGGGCTGTGTTGACCGACATGATGCTGGTGGCACGTGCCATTCGCGAGGGCGTGC 120
DB 61 GATCGGGCTGTGTTGACCGACATGATGCTGGTGGCACGTGCCATTCGCGAGGGCGTGC 120
QY 121 GACCGGGACACAGCGCTGCTGCTTTGTACAAGGGGCTGAGTCATGCGCTGGGGCT 180
DB 121 GACCGGGACACAGCGCTGCTGCTTTGTACAAGGGGCTGAGTCATGCGCTGGGGCT 180
QY 181 GGCACATAGTTGCGAAGACAGTAGTTGCGGGCAAGACAGACAGCATGAACAGCAGTCCGA 240
DB 181 GGCACATAGTTGCGAAGACAGTAGTTGCGGGCAAGACAGACAGCATGAACAGCAGTCCGA 240
QY 241 AGACGAGGAATCTCTCGGCTCTACCGGATTTGGAAGACAGACCGCTCTGAGATCGG 300
DB 241 AGACGAGGAATCTCTCGGCTCTACCGGATTTGGAAGACAGACCGCTCTGAGATCGG 300
QY 301 TTTTGTGATGTCCTGTTCCGGAGAGTCTCGAACGGGCTGTGGGGTGTCTATGCGGTC 360
DB 301 TTTTGTGATGTCCTGTTCCGGAGAGTCTCGAACGGGCTGTGGGGTGTCTATGCGGTC 360
QY 361 TGATCGGAGTGGCGAAAGTCTCTGAACAGGAGCTCCCGGGAGGCGAGTCCAGAGGTC 420
DB 361 TGATCGGAGTGGCGAAAGTCTCTGAACAGGAGCTCCCGGGAGGCGAGTCCAGAGGTC 420
QY 421 GAAATGGCTCTGTTGAAACCGTCAATCGGTTTTTACGTTTTCGGGGCTATGATGCCA 480
DB 421 GAAATGGCTCTGTTGAAACCGTCAATCGGTTTTTACGTTTTCGGGGCTATGATGCCA 480
QY 481 CATGCCCGGCTTGTGGTCCCGTCCAGCGACCGCGGCAACACAGAGAAATCCATGA 540
DB 481 CATGCCCGGCTTGTGGTCCCGTCCAGCGACCGCGGCAACACAGAGAAATCCATGA 540
QY 541 TTACGCGGGAACCCCTTAAGTCTCTTCTGCAATGTCCAGGCTCCCGCTATGACATCG 600
DB 541 TTACGCGGGAACCCCTTAAGTCTCTTCTGCAATGTCCAGGCTCCCGCTATGACATCG 600
QY 601 ACGGATCAAGCCTGGGATCGTCATTTCCGGTGTAGTAACTTTTTCGACCCCATGAGG 660

DB 601 ACGGATCAAGCCTGGGATCGTCATTTCCGGTGTAGGTAACTTTTTCGAGCCCATGAGG 660
QY 661 CGTTCTACGTGCGACAGATTTCTTAACACACGCTCCGACTGGCGCATTTGTTGGTCTTCGCC 720
DB 661 CGTTCTACGTGCGACAGATTTCTTAACACGCTCCGACTGGCGCATTTGTTGGTCTTCGCC 720
QY 721 TGACGGGACGTGACCCGTTCAAAAGAAAAGCCGAGGAATTTCAAGGCCCCAGACTGCTGT 780
DB 721 TGACGGGACGTGACCCGTTCAAAAGAAAAGCCGAGGAATTTCAAGGCCCCAGACTGCTGT 780
QY 781 ATTCCCTGACGAGACGGCTCCGTCGCGCAAGACACGGTGCCTGATGGCGCGCTGC 840
DB 781 ATTCCCTGACGAGACGGCTCCGTCGCGCAAGACACGGTGCCTGATGGCGCGCTGC 840
QY 841 GTGACTATCTGCTTCCCGCCCGCATCGGGAAGCCGCTGCTGAAGCATCTTCTTCATCCGG 900
DB 841 GTGACTATCTGCTTCCCGCCCGCATCGGGAAGCCGCTGCTGAAGCATCTTCTTCATCCGG 900
QY 901 CCATCCGCTATCTGTTCCATGACGATCAACGGAAGCGCTACAACATCAACGAGACGCG 960
DB 901 CCATCCGCTATCTGTTCCATGACGATCAACGGAAGCGCTACAACATCAACGAGACGCG 960
QY 961 GTGCGTTCGATCTCGAGAATCGGCGAGTAAAGCGCGACCTCAAGAACCCGGAAGCGGT 1020
DB 961 GTGCGTTCGATCTCGAGAATCGGCGAGTAAAGCGCGACCTCAAGAACCCGGAAGCGGT 1020
QY 1021 CTACCGTTTTCGTTACGTGCTGCGAGCCCTGCTGCTGCTGGGATGCCGCTGTAGG 1080
DB 1021 CTACCGTTTTCGTTACGTGCTGCGAGCCCTGCTGCTGCTGGGATGCCGCTGTAGG 1080
QY 1081 CATTACGGTTCATGCTCTGTGATAACCTGCTCATACGCGCAATGTCGCCCGCAAGGCT 1140
DB 1081 CATTACGGTTCATGCTCTGTGATAACCTGCTCATACGCGCAATGTCGCCCGCAAGGCT 1140
QY 1141 TCCTCGGCTATGCGAAGCGCGCATCGGAGTTGGGAGTGGGCAATGAGGAAACGCGA 1200
DB 1141 TCCTCGGCTATGCGAAGCGCGCATCGGAGTTGGGAGTGGGCAATGAGGAAACGCGA 1200
QY 1201 CTTTCCGCAACGGAATGTTGATCGCATCAACCGCGAGCTTTCGCGGAAATCCCAAGA 1260
DB 1201 CTTTCCGCAACGGAATGTTGATCGCATCAACCGCGAGCTTTCGCGGAAATCCCAAGA 1260
QY 1261 AGCTCAACGCGCGCAGTGGGCTGGATGACGACCTGCGCTGGTGGCGGAGATTTCATC 1320
DB 1261 AGCTCAACGCGCGCAGTGGGCTGGATGACGACCTGCGCTGGTGGCGGAGATTTCATC 1320
QY 1321 AGTGGTGTGGAAGACCAAGTTTGGGATGCGCGTCCGCGCTTGAANAACCGCGCTGC 1380
DB 1321 AGTGGTGTGGAAGACCAAGTTTGGGATGCGCGTCCGCGCTTGAANAACCGCGCTGC 1380
QY 1381 AGATGGTGGGACGCTGACGAGCTGGGAGTACGTCAAGATCCGATGCTCAATGCAGGC 1440
DB 1381 AGATGGTGGGACGCTGACGAGCTGGGAGTACGTCAAGATCCGATGCTCAATGCAGGC 1440
QY 1441 ATGTCTGCTGCTGCTTCCAGGCAATTCGTGCGGCTATGAGAATGTGATGACGCGATTG 1500
DB 1441 ATGTCTGCTGCTGCTTCCAGGCAATTCGTGCGGCTATGAGAATGTGATGACGCGATTG 1500
QY 1501 AAGACGCGAACTCTTGGCAATCTGAAGAACTATCTCAACAAAGATGTCATCCGAGCC 1560
DB 1501 AAGACGCGAACTCTTGGCAATCTGAAGAACTATCTCAACAAAGATGTCATCCGAGCC 1560
QY 1561 TGAAGGGGCTTTCAGGATGACGCTCGAAGGCTATCGGACAGGCTCATACGCGTTTCT 1620
DB 1561 TGAAGGGGCTTTCAGGATGACGCTCGAAGGCTATCGGACAGGCTCATACGCGTTTCT 1620
QY 1621 CCAACAAGGCGATGTGCGACACAGCTTCGCGATTTGCTAGCGATTGGTGTTCAGGTTTC 1680
DB 1621 CCAACAAGGCGATGTGCGACACAGCTTCGCGATTTGCTAGCGATTGGTGTTCAGGTTTC 1680
QY 1681 AGGTGTTCTGACGGAACCGTGGCGATTCGAGAACGAGCGGACCTGTCACGTA 1740

1681 AGGTGTTCTGGACGAAACCGTGTGCGGGCGATCGAAGACGCGGACCTGTACGTA 1740
1741 TAGCGTTGGAATTCATCTCTGAAATGCTCGTGGTGGCGACAGAGGCGGGA 1800
1741 TAGCGTTGGAATTCATCTCTGAAATGCTCGTGGTGGCGACAGAGGCGGGA 1800
1801 CGTATGAATTCGTCGAGCGGACTTATGGGAGCGCGAATGGAAGTTGGCCAAAGGCGGACG 1860
1801 CGTATGAATTCGTCGAGCGGACTTATGGGAGCGCGAATGGAAGTTGGCCAAAGGCGGACG 1860
1861 ACTTCGAAGGCTCTCTGAAGCTCCCGCGTTCGATGGTGGCGCATCTGGATAGTCCG 1920
1861 ACTTCGAAGGCTCTCTGAAGCTCCCGCGTTCGATGGTGGCGCATCTGGATAGTCCG 1920
1921 AACTGGATCAAAAGTTCATCTGCTGCGGAAGATCATCCGCGAAAGGCGTAAAGCGG 1980
1921 AACTGGATCAAAAGTTCATCTGCTGCGGAAGATCATCCGCGAAAGGCGTAAAGCGG 1980
1981 CCATCCCGCGCTGAATTCGGCTTTTAGGGTAGCGACTGAAACAGAAAAACCGGCTCTGGA 2040
1981 CCATCCCGCGCTGAATTCGGCTTTTAGGGTAGCGACTGAAACAGAAAAACCGGCTCTGGA 2040
2041 AGGACGCGGTTTTTTTATGCTCAGATCTGTCCCATCAGGACAGGATCAGGAGCCA 2100
2041 AGGACGCGGTTTTTTTATGCTCAGATCTGTCCCATCAGGACAGGATCAGGAGCCA 2100
2101 CGATCAGGACAGTCCGCTGGAGGGGAGCCCCATTTCCGAACTGTACGGCCATGACGGCA 2160
2101 CGATCAGGACAGTCCGCTGGAGGGGAGCCCCATTTCCGAACTGTACGGCCATGACGGCA 2160
2161 GCGACCCAGATCAGGATTCAGAAAGGATCAGTCCCATGGCACATCTCTTTCGCGGTT 2220
2161 GCGACCCAGATCAGGATTCAGAAAGGATCAGTCCCATGGCACATCTCTTTCGCGGTT 2220
2221 GAGACTGCTGTGTCTCCGGTGTCTAAAGTTCCTAGGGGCGGAAAGATCAAAAGC 2280
2221 GAGACTGCTGTGTCTCCGGTGTCTAAAGTTCCTAGGGGCGGAAAGATCAAAAGC 2280
2281 TGTCGGTCCGGTTAATCCGGTCCCAAGCCGATGATGCGGGCCACCCGCTCTGTCCG 2340
2281 TGTCGGTCCGGTTAATCCGGTCCCAAGCCGATGATGCGGGCCACCCGCTCTGTCCG 2340
2341 CGTTTTGCGCTGTCTGTACATAGTTTCTGGGCCAGCAGCTCCGGATGATGTCGGG 2400
2341 CGTTTTGCGCTGTCTGTACATAGTTTCTGGGCCAGCAGCTCCGGATGATGTCGGG 2400
2401 ATCAGGGTCCGACGACGCGGATTTCTGTGTGCTGCGCTGCGGGTGTATGCCGAGA 2460
2401 ATCAGGGTCCGACGACGCGGATTTCTGTGTGCTGCGCTGCGGGTGTATGCCGAGA 2460
2461 ATACGATAGGCATCCGGTCTGTTCCGCTGGCGCGGATTTGTGCCGCTTTTCGGCCCGG 2520
2461 ATACGATAGGCATCCGGTCTGTTCCGCTGGCGCGGATTTGTGCCGCTTTTCGGCCCGG 2520
2521 TCCCATGCTCTGGCGGACGCCAATGCGCCGCTGAACCGCTGCAGAAAAATCGATTCC 2580
2521 TCCCATGCTCTGGCGGACGCCAATGCGCCGCTGAACCGCTGCAGAAAAATCGATTCC 2580
2581 TTCGGGTGAAGTTCGCGGCTGGGGCGCATCGGCGACGGGCAVACGGAACAGTCCCGTC 2640
2581 TTCGGGTGAAGTTCGCGGCTGGGGCGCATCGGCGACGGGCAVACGGAACAGTCCCGTC 2640
2641 ATGAGGTTCTAAGCGGCGCGTATTATCGGCATAGGCTTTGCCATTTTCGGGCGCATAC 2700
2641 ATGAGGTTCTAAGCGGCGCGTATTATCGGCATAGGCTTTGCCATTTTCGGGCGCATAC 2700
2701 ATCTCGAATTCGTCGCTCGGTCGGGGCGGATCGAAGACAGATGCCGACTTCCTTGGTG 2760
2701 ATCTCGAATTCGTCGCTCGGTCGGGGCGGATCGAAGACAGATGCCGACTTCCTTGGTG 2760
2761 TTATCGGGGGGAACCTGGAAGCAGGCTTTGAAAGCGTTGATTTCTGTCGGTTCACCGG 2820
2761 TTATCGGGGGGAACCTGGAAGCAGGCTTTGAAAGCGTTGATTTCTGTCGGTTCACCGG 2820

QY 2821 CCGTCGATCTTTCCGAGCTTCGCGCACAGGCAACAAGCGCGATGGCGTAAAGCTGATCT 2880
Db 2821 CCGTCGATCTTTCCGAGCTTCGCGCACAGGCAACAAGCGCGATGGCGTAAAGCTGATCT 2880
QY 2881 CGTTTGGCCAGGGCGGACGAATCTTGGCAGCGCGGAAAAAGGCGCGCTGTGGGATCG 2940
Db 2881 CGTTTGGCCAGGGCGGACGAATCTTGGCAGCGCGGAAAAAGGCGCGCTGTGGGATCG 2940
QY 2941 GGAGCGCATTTCCGGGGAAGCGCTCACTCAGCGCGCCGCTTGAGGGCTTGAGTAGCGAA 3000
Db 2941 GGAGCGCATTTCCGGGGAAGCGCTCACTCAGCGCGCCGCTTGAGGGCTTGAGTAGCGAA 3000
QY 3001 CCGTTATCGCGGGATGCCCCAGCGCTCGCGCCATCAGTCTCGGAAAAAGGACCAACACC 3060
Db 3001 CCGTTATCGCGGGATGCCCCAGCGCTCGCGCCATCAGTCTCGGAAAAAGGACCAACACC 3060
QY 3061 GCGAAGCCCGGACACCAACCGAATCTTGGCCAGATAGCATGTCTCAACCTAGCAC 3120
Db 3061 GCGAAGCCCGGACACCAACCGAATCTTGGCCAGATAGCATGTCTCAACCTAGCAC 3120
QY 3121 GCCGCTCACAGCGCAATGACAGATCAGAGGCTAGGTGTAGGTGCTGATGCGCAACC 3180
Db 3121 GCCGCTCACAGCGCAATGACAGATCAGAGGCTAGGTGTAGGTGCTGATGCGCAACC 3180
QY 3181 GCCGGGCTTCGGGTGTGTAGAGCTAGGAGTTACGAACTTATCGCTGTCTCATGCTTT 3240
Db 3181 GCCGGGCTTCGGGTGTGTAGAGCTAGGAGTTACGAACTTATCGCTGTCTCATGCTTT 3240
QY 3241 TGAGGCGCAGGTTCTTCTGTTTCATGACGGAATTTTTATGCCACCTTGATCCAG 3300
Db 3241 TGAGGCGCAGGTTCTTCTGTTTCATGACGGAATTTTTATGCCACCTTGATCCAG 3300
QY 3301 ACTGCTACTTCGATCCCTTCGCTCTGATGACGAACCTGATGATCTTTGATCAAGCT 3360
Db 3301 ACTGCTACTTCGATCCCTTCGCTCTGATGACGAACCTGATGATCTTTGATCAAGCT 3360
QY 3361 CTGCAATGTGGCTGCAGAAAGTGTGAACTGGTTGCGGGAAGCGGATCAATAATGGGTT 3420
Db 3361 CTGCAATGTGGCTGCAGAAAGTGTGAACTGGTTGCGGGAAGCGGATCAATAATGGGTT 3420
QY 3421 CGGATTCGGGGGGGCTGCTTTCATGCTGGGGGGCTTCTGTCATCCTGCCGTCTCG 3480
Db 3421 CGGATTCGGGGGGGCTGCTTTCATGCTGGGGGGCTTCTGTCATCCTGCCGTCTCG 3480
QY 3481 GGTCTGTGGATGCTGCCGCTCGGCTGATTTGCTGCGCAGGATATTCCTGTTCCGT 3540
Db 3481 GGTCTGTGGATGCTGCCGCTCGGCTGATTTGCTGCGCAGGATATTCCTGTTCCGT 3540
QY 3541 GCGCTTCAGGGCGCGCTTTCGCTGGATCGAACTCAACATCCGGAATGGCTGGGCTTT 3600
Db 3541 GCGCTTCAGGGCGCGCTTTCGCTGGATCGAACTCAACATCCGGAATGGCTGGGCTTT 3600
QY 3601 CCGCGAAAAAGCGGCGAGAGCTAACCGTTGCTTGGAGCTGTTCTGAAGATGTCTAGT 3660
Db 3601 CCGCGAAAAAGCGGCGAGAGCTAACCGTTGCTTGGAGCTGTTCTGAAGATGTCTAGT 3660
QY 3661 GCTCAACCCCGCAGGGCTGAAGCCAGTGGCGCTCTGTTGCTGCGCGGATCGAGAGAA 3720
Db 3661 GCTCAACCCCGCAGGGCTGAAGCCAGTGGCGCTCTGTTGCTGCGCGGATCGAGAGAA 3720
QY 3721 GCCACAGAGACGAAAGCTCTGCTGGCGGACTGCGGCCATCGGCTCCAGATAGCCGAG 3780
Db 3721 GCCACAGAGACGAAAGCTCTGCTGGCGGACTGCGGCCATCGGCTCCAGATAGCCGAG 3780
QY 3781 AACTCGGGTTCAGTGCACGAGCTCCGCTGTCTGACAGAGACAGCTCGGTTTGACG 3840
Db 3781 AACTCGGGTTCAGTGCACGAGCTCCGCTGTCTGACAGAGACAGCTCGGTTTGACG 3840
QY 3841 AGATCACTCAATTCGGGTTGTTCTCAAGCGCTTCAAGCCCATTTGTCGGGTTTCGGGAA 3900
Db 3841 AGATCACTCAATTCGGGTTGTTCTCAAGCGCTTCAAGCCCATTTGTCGGGTTTCGGGAA 3900

QY 3901 CATCAGGTCGGATCACTCAGCAGCTCCCGCGCAGAGATATTAAGCAGCGATCGGCCG 3960
 Db 3901 CATCAGGTCGGATCACTCAGCAGCTCCCGCGCAGAGATATTAAGCAGCGATCGGCCG 3960
 QY 3961 AGTTCCCGATCCGATCAGGACAGTTACGTACGAAACCGGTTGCTCCAAATCCGTTTGACC 4020
 Db 3961 AGTTCCCGATCCGATCAGGACAGTTACGTACGAAACCGGTTGCTCCAAATCCGTTTGACC 4020
 QY 4021 GGAGAGCCAGAAAAAGCTCCGGAATGTCGCAATATCCAGCCGACACAGTTGCTCGAGT 4080
 Db 4021 GGAGAGCCAGAAAAAGCTCCGGAATGTCGCAATATCCAGCCGACACAGTTGCTCGAGT 4080
 QY 4081 TTTGGTCAATCAGCTCCGGCGGGCGCTGAAGCTT 4115
 Db 4081 TTTGGTCAATCAGCTCCGGCGGGCGCTGAAGCTT 4115

RESULT 2

AAZ35672

ID AAZ35672 standard; DNA; 1458 BP.

XX

AC AAZ35672;

XX

DT 27-JAN-2000 (first entry)

XX

DE Gluconobacter suboxydans L-sorbose reductase nucleotide sequence;

DE

KW Gluconobacter suboxydans; L-sorbose reductase; genetic engineering;

KW

KW mutant; ds.

XX

OS Gluconobacter suboxydans.

XX

FH Key Location/Qualifiers

FT 1..1458

FT /*tag= a

FT /product= "L-sorbose reductase"

XX

XX AU9920390-A.

XX

XX 23-SEP-1999.

XX

PF 11-MAR-1999; 99AU-0020390.

XX

PR 13-MAR-1998; 98EP-0104546.

XX

XX (HOFF) HOFFMANN LA ROCHE & CO AG F.

XX

XX Hoshino T, Tazoe M, Shinjoh M, Kon T;

XX

XX WPI; 1999-579276/49.

DR

DR P-PSDB; AAY49913.

XX

PT Set of genetically modified mutants not containing L sorbose reductase

XX

PS Example 4; Page 17-18; 33pp; English.

XX

CC The present invention describes a genetically engineered

CC microorganism derived from a microorganism belonging to the genus

CC Gluconobacter or Acetobacter which is characterised in that the

CC biological activity for reducing L-sorbose is substantially

CC nullified by gene recombination. The present sequence encodes

CC Gluconobacter suboxydans L-sorbose reductase.

CC N.B. This patent is equivalent to the basic NO9901197 in week 199949.

XX

SQ Sequence 1458 BP; 329 A; 386 C; 434 G; 309 T; 0 other;

XX

Query Match 23.2%; Score 955.6; DB 20; Length 1458;

Best Local Similarity 78.5%; Pred. No. 4.5e-241;

Matches 1144; Conservative 0; Mismatches 314; Indels 0; Gaps 0;

XX

QY 537 ATGATTCAGCGCGAAACCTTAAGTCTCTCTGCAATGTCCAGGCTCCCGCTATGAC 596.

Db

1 ATGATTCAGCGCGAAACCTTAAGTCTCTCTCCCGCGGTGTGCAGGCTCCCGCTATGAC 60

QY 597 ATCCAGCGGATCAAGCCTGGGATCGTGCATTTTCGGTGTAGGTAACTTTTTCGAGCCCAT 656
 Db 61 ATCAATGGGATCAAAACCGGGATCGTGCAITTTTGGCGTGGAAACTTCTTCCGGGCCCAT 120
 QY 657 GAGCGGTTCTACGTGAGCAGATTTTGAACAGCCTCCGACATGGGCGCATTTTGGTGT 716
 Db 121 GAGCGTTTCTACGTGTAACAGATCTTCAAGGAGATCCGAACTTGGGAATCATCGGCGTT 180
 QY 717 GGCTGAGCGGATGACCGTTCAAGAAAAAGCCGAGGAATTTCAAGGCCAGGACTGC 776
 Db 181 GGCTGAGCGGATGACCGTTCAAGAAAAAGCCGAGGAATTTCAAGGCCAGGACTGC 240
 QY 777 CTGTATTCTCCCTGAGCAGCGGCTCCGTCGCGCAAGAGCAGCGTCCGCTCATGGGCGG 836
 Db 241 CTCTTTTCCCTGAGCAGCGGCTCCGTCGCGCAAGAGCAGCGTTCGTGTTATGGGCGG 300
 QY 837 CTGCGTACTATCTGCTTGGCCGCGGATCCGGAAGCCGCTGCTGAAGCATCTTGTGAT 896
 Db 301 CTGAGGATTAACCTTTTGGCTCTCGCGATCCGGAAGCCGCTGCTGAAGCATCTCGCTGAC 360
 QY 897 CCGGCCATCCGATCGTTTCCATGACGATCAGGAGCGCGCTACAACATCAACGACGACG 956
 Db 361 CCGGGAATCCGATCGTTTCCATGACGATCAGGAGCGCGTTACACATTAAGGACGACG 420
 QY 957 ACCGTCGCTTGGATTCGGGAATGCGGAGTAAAGCCGACCTCAAGAACCCCGGAAG 1016
 Db 421 ACAGGTGAGTTGATCTTGAGAACAGCGGTTTCCAGCAGGATCTGAAGACACCCGAAACG 480
 QY 1017 CCGTCTACCGTTTTCGGTTAGTGGTCCGAGCCCTGCTGCTGGTGGGATCCGCTGCT 1076
 Db 481 CCGTCCCAATCTTTGGATATGTTGGAAGGATGCGCCGCGCGCTGACGAGGTGGC 540
 QY 1077 AAGCATTTACGGTCAATGCTCTGTGATAACCTGCTGATCAACGCAATGTGCGCCGCAAG 1136
 Db 541 AAGSCCTTCAGGATCATGCTCTGCGATANTCGGGCATTAACGTAATGTGCGCCGCAAG 600
 QY 1137 GCCTTCCTCGGCTATGCAAGCGCGGATCCGAGTTCGGAGTTGGGAAAGTGGATTGAGGAAC 1196
 Db 601 GCATTTCTGGGATACGCGAAGCGCGTGTCCGGAAGTGGCAAGTGGATTGAAGAGAAC 660
 QY 1197 GCGACCTTCCGACGAGCAATGTTGATCGCATCACCCGACCGTTTCGGCGGAATCGCC 1256
 Db 661 GCGAGCTTCCCAATGCGATGTTGATCGCATACGCGGACCGTTTCTGCTGACATGCG 720
 QY 1257 AAGAAGCTCAACGCGGCGAGTGGCTGATGACACCTGCGCTGGTGGCGGAGATTC 1316
 Db 721 AAGAAGCTCAACGAGCAAGTGGCTGCACGACACCTGCGCTGCTGCGAAGACTTT 780
 QY 1317 CATCAGTGGTCTGGAAGACCAAGTTTGGGATGGCGCTCCGCGCTTGAAGAACCGCGC 1376
 Db 781 CATCAGTGGTCTGGAAGACAGCTTTGCTGATGGCGGCTGCGCTGGAAGAACCGCGGA 840
 QY 1377 GTGCAGATGTTGCGGGGACGTGACGACTGGGAGTACGTCAAGATCCGAATGCTCAATGCA 1436
 Db 841 GTGCAGTTCGTTGGGATGTGACGACTACGAGCATGTAAATCCGATGCTGAATGCT 900
 QY 1437 GGGCATGTCATGCTCTCTTCCAGGATTTCTGCTGGCTATGAGATGTGGATGACGCC 1496
 Db 901 GGTACATCATGCTCTCTTCCCGGCTTCTTGGCAGGATTTGAAATGTGATCATGCC 960
 QY 1497 ATTGAACACAGCAACTCTTGGCAATCTGAGAACTATCTCAACAGGATGTCATCCG 1556
 Db 961 CTTGCTGATCCGATCTACGCGGTATCTCTGAGAACTTCTCTGAAACAGAGCTCATCCG 1020
 QY 1557 ACCCTGAAGCGCTTCAGGATGACGCTGAAAGGCTATCGGACAGCGTTCATCAGCGGT 1616
 Db 1021 ACCCTGAAGCGCTTCAGGATGACGCTGAAAGGCTATCGGACAGCGTTCATCAGCGGT 1080
 QY 1617 TTCTCCAAACAGGCGATGTCCGACGACGCTCCGAGTTGCTAGCGATGGCTGTTCGAAG 1676
 Db 1081 TTCTCGAATCCGCGCATGGCGGATCAGACATTTGCTGATTTCCCGGGGACGGAGCTCGAAG 1140

Db 7482 CGAAGAGTGGCGTACCTTACCCCGCATGTCCATGCTTACGAGCGATGAGATCCG 7541
Qy 1424 AATGCTAATCAGGCGATGCTGCTCTCCAGGCATTTCTGTCGCGCTATGAGAA 1483
Db 7542 CATCTGAACGGGGCCATGCGGTATGCGCTATCCGTCGCGCTCATGGACATCCAGCT 7601
Qy 1484 TGTGATGACGCCATTTGAAGACAGCGAACTCTTTGGCAATCTGAAGAACTATCTCAACAA 1543
Db 7602 CGTGACGCGCCATGCGCCCATCCGCTGATCGCGGCTTCTGACACAGGTCGAGGTGCA 7661
Qy 1544 GGATGCTATCCGACCCCTGAAGGCGCTTCAAGGATGACGCTCAAGGCTATCGGACAG 1603
Db 7662 GGAGATCTGCCCCATGCTCCGCGCGCTGCGGACACAGCATCCCGGACTATCTTACCCCT 7721
Qy 1604 CGTATCAGCGCTTTCTCAACAAGGCGATGTCGGACACAGCGCTCCCGGATTGCTAGCGA 1663
Db 7722 GATCAGAGCGCTTCTCGAACCCCGAGATCGCGACACAGCGCGCGCTCTGCCCTGA 7781
Qy 1664 TGGCTGTTCCAGGTTTCAGGTGTTTC 1688
Db 7782 CGGTTGGAACCGGACCGCGAAGTTTC 7806

RESULT 4

AAC90082

ID AAC90082 standard; DNA; 9810 BP.

XX

AC AAC90082;

XX

DT 19-MAR-2001 (first entry)

XX

DE AF018073 cDNA clone.

XX

KW METH; metalloprotease; thrombospondin; angiogenesis inhibition;

KW cancer therapy; benign tumour; ocular angiogenic disease;

KW rheumatoid arthritis; psoriasis; wound healing; endometriosis;

KW vasculogenesis; granulation; hypertrophic scar; nonunion fracture;

KW scleroderma, trachoma; vascular adhesion; myocardial angiogenesis;

KW coronary collateral; cerebral collateral; arteriovenous malformation;

KW ischaemic limb angiogenesis; Osler-Webber syndrome; wound granulation;

KW plaque neovascularisation; telangiectasia; haemophilic joint; EST;

KW angiofibroma; fibromuscular dysplasia; expressed sequence tag;

KW Crohn's disease; atherosclerosis; birth control; ss.

XX

Unidentified.

OS

XX

XX

PN WO200071577-A1.

XX

XX

PD 30-NOV-2000.

XX

XX

PF 25-MAY-2000; 2000WO-US14462.

XX

XX

PR 25-MAY-1999; 99US-0318208.

PR 20-JUL-1999; 99US-0144882.

PR 10-AUG-1999; 99US-0147823.

PR 13-AUG-1999; 99US-0373658.

PR 22-DEC-1999; 99US-0171503.

PR 22-FEB-2000; 2000US-0183792.

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

DR WPI; 2001-025136/03:
XX METH1 and METH2 polynucleotides and encoded polypeptides, used to
PT inhibit angiogenesis in the treatment of disorders such as cancer,
PT rheumatoid arthritis and psoriasis
XX
PS Claim 7; Pages 653-659; 768pp; English.
XX
CC The present invention relates to human METH1 and METH2, (ME for
CC metalloprotease and TH for thrombospondin; see AAB50002 and AAB50003).
CC The present sequence is an expressed sequence tag (EST) for METH. METH
CC can be used for inhibiting angiogenesis in an individual, and for
CC treating cancer, benign tumours, an ocular angiogenic disease,
CC rheumatoid arthritis, psoriasis, delayed wound healing, endometriosis,
CC vasculogenesis, granulations, hypertrophic scars, nonunion fractures,
CC scleroderma, trachoma, vascular adhesions, myocardial angiogenesis,
CC coronary collaterals, cerebral collaterals, arteriovenous malformations,
CC ischaemic limb angiogenesis, Osler-Webber syndrome, plaque
CC neovascularisation, telangiectasia, haemophilic joints, angiofibroma,
CC fibromuscular dysplasia, wound granulation, Crohn's disease or
CC atherosclerosis. METH can also be used in birth control. METH can also
CC be used in diagnostic methods for the prognosis of cancer.
XX
SQ Sequence 9810 BP; 1583 A; 3401 C; 3201 G; 1625 T; 0 other;

Query Match 6.0%; Score 247.4; DB 22; Length 9810;
Best Local Similarity 53.4%; Pred. No. 2.8e-54;
Matches 590; Conservative 0; Mismatches 506; Indels 9; Gaps 3;

Qy 584 TCCCCCTATGACATCGACGGGATCAAGCTGGGATTCGATTTTCGGTGTAGTAACCT 643
Db 6711 TCCCTCTATACCGCAAGGCGCTCACTCCCGGCATGTCATATCGGCTCGGCACTT 5770
Qy 644 TTTTCGAGCCCATGAGGCGTTTACGTCGAGCAGATTTTGAACACGCTCCGACTGGGC 703
Db 6771 CCACGGGGCATGACGGGCTTATCTCGAGCATCTTTCGGCTGGGCGGACGACGA 6830
Qy 704 GATTGTTGGTGTGGCCTGACGGGAGTGACGCTTCAAGAAAAAGCGGAGGATTCA 763
Db 6831 CTGGGCCATCTCTCGGCGGGCGTCCGCGGACCATCGCGGATGCGGAGGCTCTGGC 6890
Qy 764 GCGCCAGGACTGCTGTATTCCTGACGAGAGCGCTCGTCCGCGAGAGCAGCGTGGC 823
Db 6891 CGCGAGGACATCTCTCGAGCATGTCGAGCTCGATCCGGCGGCGCA---CCGGCGCG 6947
Qy 824 CGTCATGGGCGCGTGGCTGACTATCTGTTCCCGCGCGCGATCGGAAAGCGCTGAA 883
Db 6948 GCAGGTGGGGCGATGCTGGGCTTCTCCGCTCGAGGCGGCAATGCGGCGCTGATCGA 7007
Qy 884 GCATCTTGTGATCGGCGCATCCGCTGTTTCCATGAGCATCAGGAAGCGGCTACAA 943
Db 7008 GGCCATGTCGGATCCGCGCATCCGCTGCTGCTGACCGTACCGAGGCGGCTATT 7067
Qy 944 CATCAACGAGACGCGGTGCTTCTGATCTGGAGAAATGCGGAGTAAAGCGGACCTCAA 1003
Db 7068 TGTGATGCTCTG---GGGCGCTTCGATCCGAGCATCCGATATGTCGCGGATCGCGC 7124
Qy 1004 GAACCCGGAAGGCGCTTACCGTTTTCGTTTACGTTGTCGAGGCGCTCGCTGCTGTTG 1063
Db 7125 CCATCTGCGCGCGCGGCGGCTTTCGCGCGGATCTCTCGGCGGCTCGGCGCGCGC 7184
Qy 1064 GGATGCGGCTGTAAGGCAATTTACGCTCATGTCCTGATACCTGCGTATACGCGCAA 1123
Db 7185 CCACGCGGGGTTTACACCTTTCACCGTGTATGCTCGGACACCTCCCCGGAACGCGCA 7244
Qy 1124 TGTGCGCCGCAAGGCTTCTCTCGGTATGCGAAGGCGCGGATCGCGAGTTGGCGAAGTG 1183
Db 7245 TGTACCCGCAACGCGGTGTTGGGCTTGGCGGAGCTCTACGAGCGGAGCTTGGCGGCTG 7304
Qy 1184 GATTGAGGAAACGCGACCTTCCCGAACGGAATGTTGATCGATCATACCCCGACCGTTTC 1243
Db 7305 GGTGAAGCGGAGGTGGCGCTTCCGAAACGCGATGTTGTCGACCGCATCACCCCGCCCGG 7364

Db 830 AGGATGCTTCAACCGAGGCGCCCGCTACGAGGAGTTGGCGTCAGGTCGTCCTCCG 889
Qy 1393 ACSTGAGGAGTGGAGTCAAGATCGAATGCTCAATGCGAGGCGATGTCATGCTCT 1452
Db 890 ACGTGGAGCCTTATGAATTAATGAAGTGGCCCTGCTCAAGCGCTCCACCGAGGACTTT 949
Qy 1453 GCTTCCCGAGGATCTGCTGCGCTATGAGATGTGATGAGCGCATGAGACAGCGAAC 1512
Db 950 GCTACTTCGGCCACTTGGCTGGCCACCACTGTTGCCAGGAGTTCATGCGGATACCCGCT 1009
Qy 1513 TCCTTGGCACTGAAGAACTATCTCAACAGGATGTCATCCCGACCTGAAGCGGCTT 1572
Db 1010 TCCAGGATTTCTCTGCTTACATGAGCGCGAAGCCACCCCTACCTCAAGGAACCTC 1069
Qy 1573 CAGCATGACGCTCGAAGGCTATCGGACACGCTATCGCGCTTCTTCCAAACAGCGA 1632
Db 1070 CAGGTGTCGATCTAGATGCTTATCGACGCCAACTATCGCGCGATTCGGCAACCCGAG 1129
Qy 1633 TGTCGGACACAGCTCCGGATGCTAGCGATGCTGTTCCAAAGGTTTCAGGTTCGGA 1692
Db 1130 TCAAAGACACCGTACCGCGCTGTGTGCGGAATCTCCGACCGCATTCCAAAGTGGCTGT 1189
Qy 1693 CGGAACGCTGCTGGCGGATCGAAGACAGCGGACCTGTCACGTATACCGTTTCGGA 1752
Db 1190 TGCCAGTGTACGCGAAACCTCGCAGCAGCGCGAGCTCACACTTCTCGACCCATCG 1249
Qy 1753 TTGCATCTATCTCGAATGCTGCTGTCGCGACGAGAGGCG 1796
Db 1250 TCGCATCTGGCGGCTACCGACAGGACCGACCGAGCAGGCG 1293

RESULT 6

AAAF71527

ID AAAF71527 standard; DNA; 1632 BP.

XX AC AAAF71527;

XX AC AAAF71527;

XX 30-APR-2001 (first entry)

XX Corynebacterium glutamicum SMP protein nucleotide sequence SEQ ID NO:335.

XX Corynebacterium glutamicum; carbon metabolism and energy production;
KW SMP protein; sugar metabolism and oxidative phosphorylation protein;
KW fine chemical production; organic acid; proteinogenic amino acid;
KW nonproteinogenic amino acid; purine base; pyrimidine base; nucleoside;
KW nucleotide; lipid; saturated fatty acid; unsaturated fatty acid; diol;
KW carbohydrate; aromatic compound; vitamin; cofactor; polyketide; enzyme;
KW diagnosis; Corynebacterium diphtheriae; evolutionary study; ds.

XX Corynebacterium glutamicum.

XX OS WO200100844-A2.

XX PN 04-JAN-2001.

XX PD 23-JUN-2000; 2000WO-IB00943.

XX PF 25-JUN-1999; 99US-0141031.

XX PR 08-JUL-1999; 99DE-1031412.

XX PR 08-JUL-1999; 99DE-1031413.

XX PR 08-JUL-1999; 99DE-1031419.

XX PR 08-JUL-1999; 99DE-1031420.

XX PR 08-JUL-1999; 99DE-1031424.

XX PR 08-JUL-1999; 99DE-1031428.

XX PR 08-JUL-1999; 99DE-1031431.

XX PR 08-JUL-1999; 99DE-1031433.

XX PR 08-JUL-1999; 99DE-1031434.

XX PR 08-JUL-1999; 99DE-1031510.

XX PR 08-JUL-1999; 99DE-1031562.

XX PR 08-JUL-1999; 99DE-1031634.

XX PR 09-JUL-1999; 99DE-1032180.

XX PR 09-JUL-1999; 99DE-1032227.

PR 09-JUL-1999; 99DE-1032230.
PR 09-JUL-1999; 99US-0143208.
PR 14-JUL-1999; 99DE-1032924.
PR 14-JUL-1999; 99DE-1032973.
PR 14-JUL-1999; 99DE-1033005.
PR 27-AUG-1999; 99DE-1040765.
PR 31-AUG-1999; 99US-0151572.
PR 03-SEP-1999; 99DE-1042076.
PR 03-SEP-1999; 99DE-1042086.
PR 03-SEP-1999; 99DE-1042087.
PR 03-SEP-1999; 99DE-1042088.
PR 03-SEP-1999; 99DE-1042095.
PR 03-SEP-1999; 99DE-1042123.
PR 03-SEP-1999; 99DE-1042125.
XX (BADI) BASF AG.
XX Pompejus M, Kroeger B, Schroeder H, Zelder O, Haberhauer G;
PI WPI; 2001-061975/07.
XX P-PSDB; AAB79410.
XX New isolated Corynebacterium glutamicum nucleic acid encoding a sugar
PT metabolism and oxidative phosphorylation protein for production or
PT modulation of production of fine chemicals e.g. amino acids,
PT carbohydrates or enzymes
XX Claim 3; Page 613-615; 1246pp; English.
XX AAAF71360 to AAAF71750 encode the Corynebacterium glutamicum sugar
CC metabolism and oxidative phosphorylation (SMP) proteins given in
CC AAAF79243 to AAB 79633 which are involved in carbon metabolism and
CC energy production. The C. glutamicum SMP gene can be used in vectors
CC (II) for expression in host cells and production or modulation of
CC production of fine chemicals, such as, an organic acid, a proteinogenic
CC or nonproteinogenic amino acid (preferred), a purine or pyrimidine base,
CC a nucleoside, a nucleotide, a lipid, a saturated or unsaturated fatty
CC acid, a diol, a carbohydrate, an aromatic compound, a vitamin, a
CC cofactor, a polyketide, or an enzyme. The presence of (I) or SMP proteins
CC (III) encoded by them are used for diagnosing the presence of (I) or host cells
CC Corynebacterium diphtheriae in a subject. (I), (II), (III) or host cells
CC containing them are used to map genomes of organisms related to
CC C. glutamicum, identify and localise C. glutamicum sequences of interest,
CC in evolutionary studies, in determining SMP protein regions required
CC for function, in modulating SMP protein activity, in modulating the
CC metabolism of sugars, and in modulating high-energy molecule production
CC in a cell (i.e. ATP, NADPH).
XX Sequence 1632 BP: 374 A; 509 C; 400 G; 349 T; 0 other;
SQ Query Match 5.2%; Score 213.6; DB 22; Length 1632;
Best Local Similarity 49.6%; Pred. No. 1e-45;
Matches 617; Conservative 0; Mismatches 609; Indels 18; Gaps 2;
Qy 568 CTGCCAATGTCAGGCTCCCGCTATGACATCGACGGGATCAAGCTGGGATCGTCATT 627
Db 153 CTTCGGAGTGCAGATCCAGCTTCAACCGCTGACGCTCGCCCGGCAATGTCCACT 212
Qy 628 TCGGTGTAGTAACTTTTTCGAGCCCATGAGGCGTTCAGTCGAGCAGATTCTTGAAC 687
Db 213 TCGGTGTGCGGATTCATCGCTCACCAAGGATGATGCTCAATGAATGATGAATG 272
Qy 688 ACCTCGGAGCTGGCGGATTTGTTGGTGTGCTGACGGGAGTACCGTTCAAGAAAA 747
Db 273 AGGCAAGGCTTGGATTGGGCGCATCGCGATGGGTGTCATGCTTCCGATGTGCCA 332
Qy 748 AAGCCGAGGAATCAAGCCCGCAGGACTGCCTGTATTCCTCGACGAGACGCTCGGTCG 807
Db 333 TGCAGATGCTCGCCAGCAAGATACCTTTTATACCTCGACCACTAAAGCTCTCTGATG 392
Qy 808 CCAAGACCGTGGCGCTCATGGCGCTGCGGTGCTGCTGCTGCTGCTGCTGCTGCTG 867

Db 393 GAACCTCTTGATCAAAAAATCATCGGATCCATCATGACTACGTGTCTCCCGAGGACC 452
QY 868 CGGAAGCCGTGCTGAAGCATCTTGTGTATCGGGCCATCCGCATCGCTTTCCATGACGATCA 927
Db 453 CAGCAGCGCGCTTGCACACCTTCGGCAGGACTCCATCCGATTTTCCCTCAGGTA 512
QY 928 CGGAAGCGCGCTTACAACTATCAACGAGACGACCGGTGCTTGTGATCTGGAGATCGGGCAG 987
Db 513 CTGAAGCGCGGATACAACTATCGATCGCGCAGAGAGATTCGACACACCAACCTCGAA 572
QY 988 TAAAGCGCGACCTCAAGAACCGGAAAGCGCTCT-----ACCGTTTCG 1032
Db 573 TCGTTGCTGACCGGAACCTCTGACGGCGGGATFACHTCCACTTTCGAGACCTTCTTTG 632
QY 1033 GTTACGTGCTGAGGCGCTTGGCTGCTGTTGGGATGCGCGTGGTAAGGCATTTACGGTCA 1092
Db 633 GTTGTATCACTGCCGCAATGATTTCCCGAAAGAAATCAGGATCTACGCCATTTACCATCA 692
QY 1093 TGTCTGTGATAACCTGCTCATACGCGCATGTGCGCCGGAAGCCCTTCCTCGGCTATG 1152
Db 693 TGAGCTGCGTAACATCCAAAGCAACGGCGATCTGGCTTAAGCGTTTCTTCCTCGCCTTCG 752
QY 1153 CGAAGCGCGCGATCCGAGTTGGCGAAGTGGATTGAGGAAACGCGACCTTCCCGAACG 1212
Db 753 CACATTCGCTGCTCTGAGCTCGCGGAATGGTGGAAACAGCTGGCCTTCCCAACT 812
QY 1213 GAATGGTTGATCGCATACCCCGACCGCTTTCGGCGGAAATCGCCAAAGACTCAACGCG 1272
Db 813 CCATGTGGACCGCATACCC---CTGAACACCGGACCGCGGATGACATCAAG 869
QY 1273 CCAGTGGGCTGGATGACACACCTCGCGGTGGCGGAGGATTTCCATCAGTGGGTCTGG 1332
Db 870 AAATCGGCTACATCGATCGTGGCGAGTGGTTCTGAAGATTTCAACCAATGGGTCTCG 929
QY 1333 AAGACCACTTTGCGGATCGCGCTCGCGCTTGAAGAACGCGGTGCGAGATGTCGGGG 1392
Db 930 AGGATGCTTCAACCGGCGCGCGCGCTACGAGGAGTGGGTGCGAGCTGCTCTCCG 989
QY 1393 ACGTGACGACTGGGAGTACGATCAAGATCCGAATGCTCAATGCGAGGATGCTCATGCTCT 1452
Db 990 ACGTGGAGCCTTATGAATTAATGAAGTGCCTGCTCAACGCTCCACACGAGGACTT 1049
QY 1453 CTTTCCCGAGGATCTGCTGGCTATGAGATGAGATGAGATGAGATGAGATGAGATGAGAT 1512
Db 1050 GCTACTTCGGGCACTTGGCTGGCCACACATGGTCCAGAGCTGATGGCGGATACCCGCT 1109
QY 1513 TCCTTGGCAATCTGAAGAACTATCTCAACAGGATGTCATCCGACCTGAAGCGCTT 1572
Db 1110 TCCAGGATTTCTCTGCTTACATGGAGCGGAGCCACCCCTACCTCAAGAACTC 1169
QY 1573 CAGGATGACGCTCGAAGGCTATCGGGACACGCTCATCAGCCGTTTCTTCCAAAGCGGA 1632
Db 1170 CAGGTGTGATCTAGATCTTATGACGCCAACTCATCGCGGATTCGGCAACGCGGAG 1229
QY 1633 TGTCCGACACGCTCCGGATGCTACGATGCTTCAAGGTTTCAGGTTCGGA 1692
Db 1230 TCAAGACACCGTACCGCGCTGTGCGGAATCTCCGACCGGATTCGAAGTGGCTGT 1289
QY 1693 CGGAACCGCTGCTGGCGGATCAAGAACAGCGGACCTGTACAGTATACGCTTCGGAA 1752
Db 1290 TGCCAGTCGTACGCGAAGAACCTCGCAGCAGCGCGGAGTCACTTCTTCGACCCATCG 1349
QY 1753 TTGATCTCTATCTCGAATGCTGGTGTGTCGCGAGGAGGGC 1796
Db 1350 TCGCATCTCTGGCGCGCTACGCAAGGACCGGACCGGAGCGGC 1393

RESULT 7

AAH64966/c

ID AAH64966 standard; DNA; 349980 BP.

XX

AC AAH64966;

XX

DT 26-SEP-2001 (first entry)

XX C glutamicum coding sequence fragment SEQ ID NO: 1.

XX Coryneform bacterium; amino acid synthesis; vitamin; saccharide;

KW organic acid synthesis; ds.

XX Corynebacterium glutamicum.

OS EP1108790-A2.

PN 20-JUN-2001.

PD 18-DEC-2000; 2000EP-0127688.

XX 16-DEC-1999; 99JP-0377484.

PF 07-APR-2000; 2000JP-0159162.

PR 03-AUG-2000; 2000JP-0280988.

XX (KYOW) KYOWA HAKKO KOGYO KK.

PA Nakagawa S, Mizoguchi H, Ando S, Hayashi M, Ochiai K, Yokoi H;

PI Tateishi N, Senoh A, Ikeda M, Ozaki A;

XX WPI: 2001-376931/40.

DR Novel polynucleotides derived from Coryneform bacteria, for identifying

XX mutation point of a gene, measuring expression of a gene, analysing

PT expression profile of pattern of a gene and identifying homologous gene

PT Claim 7; SEQ ID NO: 1; 246pp + Sequence Listing; English.

XX The present invention provides a number of nucleotide and protein

PS sequences from the Coryneform bacterium Corynebacterium glutamicum. These

CC are useful for identifying the mutation point of a gene derived from a

CC mutant of coryneform bacterium, measuring expression amount and

CC analysing the expression profile or expression pattern of a gene derived

CC from Coryneform bacterium, and identifying a homologue of a gene derived

CC from coryneform bacterium. Coryneform bacteria are useful for producing

CC amino acids, nucleic acids, vitamins, saccharides and organic acids,

CC particularly L-lysine. The present sequence is a nucleic acid described

CC in the exemplification of the invention.

CC Note: The sequence data for this patent did not form part of the printed

CC specification, but was obtained in electronic format directly from the

CC European Patent Office.

XX Sequence 349980 BP; 79703 A; 91547 C; 98381 G; 80349 T; 0 other;

SQ Query Match 5.2%; Score 213.6; DB 22; Length 349980;

Best Local Similarity 49.6%; Pred. No. 1.1e-44;

Matches 617; Conservative 0; Mismatches 609; Indels 18; Gaps 2;

QY 568 CTGCGCAATGTCAGGCTCCCGCTTATGACATCGACGGGATCAAGCTTGGGATCGCATT 627

Db 122407 CTTCCGGAGTGCATCCCGCTTCAACCGCGTACGCTCGCGCGGCTTGTCCACT 122348

QY 628 TCGGTGTAGTAACTTTTTCAGCCCATGAGGCGTCTACGTCGACGAGATTTCTTGAAC 687

Db 122347 TCGGTGTGGCGGATTCATCGCGCTCACCAGCGATGTACTCAATGAATGATGATG 122288

QY 688 ACGTCCGGACTGGCGGATTTGTTGGTGTGGCTTGACGGGCGAGTACCGTTCAAGAAAA 747

Db 122287 AGGCAAGGCCCTTGGATTGGGGCATCATCGCGATGGTGTCTCCGATGTGCGCA 122228

QY 748 AAGCCGAGGAATTCAGGCCCGGAGGCTCCCTGATTCCTACCGAGAGGCTCCGTCG 807

Db 122227 TGGCGATGCGCTGCGCGCTCATGGCGCGCTGCGGTACTATCTGCTGCCCGCGGATC 122168

QY 808 GCAAGAGCACGGTGCCTGATGGCGCGCTGCGGTACTATCTGCTGCCCGCGGATC 867

Db 122167 GAATCTTGATCAAAAAATCATCGGATCCATCATGACTGCTTCCGCTCCCGAGGACC 122108


```
Db      248 TCGACCGCATGTTCCGGCTGCAACCGACGAA 217
RESULT 10
AAD16810
ID      AAD16810 standard; DNA; 1848 BP.
XX
AC      AAD16810;
XX
XX      29-NOV-2001 (first entry)
XX
DE      Escherichia coli strain C arabinol dehydrogenase gene.
XX
XX      Positive selection system; metabolise; arabinol; ribitol; mannitol;
KW      transgenic cell; marker gene; arabinol dehydrogenase; ds.
XX
XX      Escherichia coli C.
OS
XX      WO200166779-A2.
XX
XX      13-SEP-2001.
XX
XX      08-MAR-2001; 2001WO-US07474.
PF
XX
XX      08-MAR-2000; 2000US-0188291.
PR
XX      15-AUG-2000; 2000US-0255595.
XX
XX      (UYGE-) UNIV GEORGIA RES FOUND INC.
PA
XX
XX      Parrott W, Lafayette P, Kane P;
PI
XX      WPI; 2001-565596/63.
DR
XX
XX      Positively selecting transformed cells comprising selectable marker
PT      gene and desired gene, from a cell population by using marker compounds
PT      e.g., arabinol, ribitol which confer selective advantage on transformed
PT      cells
XX
XX      Claim 1; Page 34-35; 37pp; English.
XX
XX      The present invention relates to a positive selection system that
CC      involves conferring to transformed cells the ability to metabolise
CC      arabinol, ribitol and/or mannitol. The positive selection method is
CC      used in positively selecting transgenic cells from a population of cells
CC      using the positive selection method, the presence of the gene of
CC      interest in the genetically transformed cells may be determined without
CC      the disadvantages associated with traditional negative selection
CC      systems. Positive selection of the transformed cells is achieved without
CC      directly damaging the neighbouring non-transformed cells. The
CC      transformed cells may be identified by simple visual means without the
CC      use of a separate assay to determine the presence of a marker gene. This
CC      technique also avoids the release of antibiotics or other dangerous
CC      genes into the environment. The present sequence is Escherichia coli
CC      strain C arabinol dehydrogenase gene.
XX
XX      Sequence 1848 BP; 494 A; 443 C; 473 G; 438 T; 0 other;
SQ
Query Match      2.6%; Score 105; DB 22; Length 1848;
Best Local Similarity 46.8%; Pred. No. 4, 1e-17;
Matches 509; Conservative 0; Mismatches 560; Indels 18; Gaps 5;
QY      622 TCGATTCGGTGTAGTAACATTTTCGAGCCCATGAGCGCTTCTACGTCGAGCAGATTC 681
Db      1111 1111 1111 1111 1111 1111 1111 1111 1111 1111 1111 1111
QY      503 TGCACATCGGTTAGTGTCTTTTCATCGCGCATACAGCGGTGGTATCTACACCGTTTGC 562
Db      1111 1111 1111 1111 1111 1111 1111 1111 1111 1111 1111 1111
QY      682 TTGACACCGCTCCGACTGGCGATGTTGGTGTGGCTGACGGCAGTGGACCTTCAA 741
Db      1111 1111 1111 1111 1111 1111 1111 1111 1111 1111 1111 1111
QY      563 AGGTGATGGCGGATAAAGCGCTGCGGCAATATCGTAATGA---TGTG 619
Db      1111 1111 1111 1111 1111 1111 1111 1111 1111 1111 1111 1111
QY      742 AGAAAAAGCCGAGGAATTCAGGCGCCAGGACTGCTGTATTCCCTGACCGAGCGGCTC 801
Db      1111 1111 1111 1111 1111 1111 1111 1111 1111 1111 1111 1111
QY      620 ACATGTCTGACAGGCATCAGTGGCACAGAAAGTTCGTATGTGCTGGAAACCGTCAGCC 679
Db      1111 1111 1111 1111 1111 1111 1111 1111 1111 1111 1111 1111
QY      802 CGTCCGCAAGACACCGTGCCTCATGGCGCGCTGCGTGAGCTATCTGCTTCCCGCG 861
Db      1111 1111 1111 1111 1111 1111 1111 1111 1111 1111 1111 1111
QY      680 CGGAAGGGTAAGCGAATATGAGAGATCACTCAATTGAGAGTTGATACCGTGGCAGG 739
Db      1111 1111 1111 1111 1111 1111 1111 1111 1111 1111 1111 1111
QY      862 CCGATCCGGAAGCGTCTGNAAGCATCTTGTGATCCGGCCATCCGATCGTTTCCATGA 921
Db      1111 1111 1111 1111 1111 1111 1111 1111 1111 1111 1111 1111
QY      740 CAGATTTACAAACCGCTGATTGCTGAAGGGCAGATCCGAGACAAAGTAGTATTCTTTCA 799
Db      1111 1111 1111 1111 1111 1111 1111 1111 1111 1111 1111 1111
QY      922 CGATCACGGAAGCGGCTTACAACATCAACGAGACGCGGTGCGTTGATCTGAGAGATG 981
Db      1111 1111 1111 1111 1111 1111 1111 1111 1111 1111 1111 1111
QY      800 CCGTACCGGAAGCGGCTACTACCTGA---ATACGAGTACAAACTGGAAGTAAACAATC 856
Db      1111 1111 1111 1111 1111 1111 1111 1111 1111 1111 1111 1111
QY      982 CGGAGTAAGAGCGGCGCTCAAGAACCGGAAAGCGCTCTACCGTTTCCGGTTACGTGG 1041
Db      1111 1111 1111 1111 1111 1111 1111 1111 1111 1111 1111 1111
QY      857 CTGATTTAGCGGCAGATCTTTAAGGGGATGCAAAACAATTTACGGTGTATTACCCGTA 916
Db      1111 1111 1111 1111 1111 1111 1111 1111 1111 1111 1111 1111
QY      1042 TCGAGGCCCTGCTGCTGTTGGGATCGCGGTGAAGGCAATTTACGGTCTATGCTCTGTG 1101
Db      1111 1111 1111 1111 1111 1111 1111 1111 1111 1111 1111 1111
QY      917 TCCTCGAAGCGGTATGGCAAAATACGCCGG-----ACCACTAACCTGCTGAATTGCG 970
Db      1111 1111 1111 1111 1111 1111 1111 1111 1111 1111 1111 1111
QY      1102 ATAACCTGCTCATACGCGCAATGTGCGCCCAAGGCTTCTCTCGGCTATCGAAG---G 1158
Db      1111 1111 1111 1111 1111 1111 1111 1111 1111 1111 1111 1111
QY      971 ATAAGTGGCCCATATGTTGTAACGTTTCCATGATGGCTGCTGAGTTTCTCAGCTAA 1030
Db      1111 1111 1111 1111 1111 1111 1111 1111 1111 1111 1111 1111
QY      1159 GCGCGATCCGAGTTGGCGAAGTGGATTGAGAAACCGGACCGACCTTCCCGAACGGAATGG 1218
Db      1111 1111 1111 1111 1111 1111 1111 1111 1111 1111 1111 1111
QY      1031 CTGGCAACAGGATGTCATCGACTGGCTGAGTACAAATACCACTTGCCTGCAATACCATGG 1090
Db      1111 1111 1111 1111 1111 1111 1111 1111 1111 1111 1111 1111
QY      1219 TTGATCGCATCACCCCGACCGTTTTCGGCGGAAATCGCCAAAGCTCAACGCGGCACTG 1278
Db      1111 1111 1111 1111 1111 1111 1111 1111 1111 1111 1111 1111
QY      1091 TTGACCGCATTACCGCTGCTCCGCGAGCAGAACTTCCGGCAGGATCAAGGCTCAACGCG 1150
Db      1111 1111 1111 1111 1111 1111 1111 1111 1111 1111 1111 1111
QY      1279 GGCTGGATGACGACTGCGCTGGTGGCCGAGGATTTCCATCAGTGGGTGCTGGAAGACC 1338
Db      1111 1111 1111 1111 1111 1111 1111 1111 1111 1111 1111 1111
QY      1151 GTATTGCGATAAAGCGCGGTAAATGGCGCAAACTTTATCCAGTGGGTGCTGGAAGATA 1210
Db      1111 1111 1111 1111 1111 1111 1111 1111 1111 1111 1111 1111
QY      1339 AGTTTGGGATGGCGCTCCGCGCTTGAAGGCGGCGTGCAGATGCTGGGACGCTGA 1398
Db      1111 1111 1111 1111 1111 1111 1111 1111 1111 1111 1111 1111
QY      1211 ATTTCCGCTGATGTCGCTCCGCGCTGGGAGAGTGGTGTGCACTGGTGGCGTGGTAA 1270
Db      1111 1111 1111 1111 1111 1111 1111 1111 1111 1111 1111 1111
QY      1399 CGGACTGGGAGTACGTCGAAGATCCGAATGCTCAATGCGAGGCGATGCTGCTGCTTCC 1458
Db      1111 1111 1111 1111 1111 1111 1111 1111 1111 1111 1111 1111
QY      1271 TCCCTATGAGAGCGGGAAGATTCGCAATCTTAACCTTTCACACAGTTGCTGCTGGG 1330
Db      1111 1111 1111 1111 1111 1111 1111 1111 1111 1111 1111 1111
QY      1459 CAGGATTTCTGGTGGCTATGAGATGAGATGGAGCCCATTTGAGACAGCGCACTCTG 1518
Db      1111 1111 1111 1111 1111 1111 1111 1111 1111 1111 1111 1111
QY      1331 CAGGTACGTTAATCGGTCAAAATATATCCAGAAAGCACAATGACCGATTTTATCTATC 1390
Db      1111 1111 1111 1111 1111 1111 1111 1111 1111 1111 1111 1111
QY      1519 GCAATCTCAAGAACTATCTCAACAGGATGTCATCCCGACCGCTGAAGCGCTTCCAGCA 1578
Db      1111 1111 1111 1111 1111 1111 1111 1111 1111 1111 1111 1111
QY      1391 AGATTGGCGCGCTACGTGACGGAAGATGTCATTCCTTGCTT---GGCGGATACGGTA 1447
Db      1111 1111 1111 1111 1111 1111 1111 1111 1111 1111 1111 1111
QY      1579 TGACGCTCGAAGGCTATCGGGACAGCGCTCATCAGCGGTTTCTCCAAAGCGGATGTCGG 1638
Db      1111 1111 1111 1111 1111 1111 1111 1111 1111 1111 1111 1111
QY      1448 TCGATTTGCCAACCTACCGTGTGTTGCTACCAACGCTTTTACCAATCCACATATTCAGG 1507
Db      1111 1111 1111 1111 1111 1111 1111 1111 1111 1111 1111 1111
QY      1639 ACCAGACGCTCCGGATTCGTAGCGTGGCTGTTTCCAAAGGTTTCAGGTGTTTCGACGGAA 1698
Db      1111 1111 1111 1111 1111 1111 1111 1111 1111 1111 1111 1111
QY      1508 ACACCAACCAACGCGCTCGGATGGTTTCTCGAAAAATTCGCGGATGATTTCGCCGCCCA 1567
Db      1111 1111 1111 1111 1111 1111 1111 1111 1111 1111 1111 1111
QY      1699 CCGTGGC 1705
Db      1111 1111 1111 1111 1111 1111 1111 1111 1111 1111 1111 1111
QY      1568 CACTGCG 1574
Db      1111 1111 1111 1111 1111 1111 1111 1111 1111 1111 1111 1111
RESULT 11
AAFI2753
ID      AAFI2753 standard; cDNA; 684 BP.
XX
XX      AAFI2753;
XX
XX      13-MAR-2001 (first entry)
XX
```


17 GAACCTTTCCGCGCTGGTAAATAGACATAAATTTGTCGCGACGCTCCGGAATGGAA 76
1368 AAAGCCGCGTGCAGATGGTCCGGGAGCTGAGGAGTACGTCAGATCCGATG 1427
77 AAAGCGGAGCGGAACTGGTGGATGATGCTGCTTATGAAGAGATGAAGTTGCGCATG 136
1428 CTCATGAGGCGCATGTCATGCTGCTTCCAGGCAATTCGTGCGGTATGGAATGTG 1487
137 CTCACGCGCATGCTTCTGCGGTATCTGGGTATCTTGAGGATATCASCACAT 196
1488 GATGAGCGCATGTAAGACAGCAACTCTCTGGCAATCTGAAGAACTATCTCAACAAGGAT 1547
197 AATGACTGTATGAAGATGAACATATATCGTTATGCGGCGTATGCTTGCATCTTCAGGAA 256
1548 GTCATCCGCGCTGAGGCGCTTCAGGCGATGAGCTGCAAGGCTATCGGAGCAGCGTC 1607
257 CAAGCGCGGAGCTTGAAGTGC---AGGGCGTTGATTTGCAAGATTAGCTTAACCGATTA 313
1608 ATCAGCGCTTCTCCAAAGCGGATGTCGACAGAGCGCTCCGATTGCTAGCGATGGC 1667
314 ATTGACGCTATAGCAACCGCGGTAGCTCATGCAACCTGGCAGATTGCGATGGATGGT 373
1668 TGTTCGAAGGTGAGGTCTTTCGACGAAACCGTGGTGGCGGATCGGAAGCAAGCGG 1727
374 AGCCAGAAATGCCACAGCGGATGTTGATTTCTGTCGCTGCGATCTGGCGCATGACAGC 433
1728 GACCTGTACGCTATAGCGTTCGGAATTCGCAATCTCATCTCGAAATGCTCGGTGGCGAC 1787
434 AAGTTCGATCTGTCGGCGCTGGCGTGGGTTGGATGCGTTATGTCGGTGGTGTGAT 493
1788 GAGAAGCGCGGACGCTATGAA 1808
494 GAACAGGGAATCCGATGAA 514

RESULT 13

AAAF08165
ID AAF08165 standard; cDNA: 621 BP.
XX
AC AAF08165;
XX
DT 13-MAR-2001 (first entry)
XX
DE Fusarium venenatum EST SEQ ID NO:688.
XX
KW Multiple gene expression; filamentous fungal cell; EST;
KW expressed sequence tag; Fusarium venenatum; Aspergillus niger;
KW Aspergillus oryzae; Trichoderma reesei; identification; recombination;
KW culture condition; environmental stress; spore morphogenesis;
KW metabolic pathway engineering; catabolic pathway engineering; ss.
XX
OS Fusarium venenatum.
XX
PN WO200056762-A2.
XX
PD 28-SEP-2000.
XX
PF 22-MAR-2000; 2000WO-US07781.
XX
PR 22-MAR-1999; 99US-0273623.
XX
PA (NOVO) NOVO NORDISK BIOTECH INC.
XX (NOVO) NOVO NORDISK AS.
XX
PI Berka RW, Rey MW, Shuster JR, Kauppinen S, Clausen IG, Olsen PB;
XX WPI: 2000-594572/56.
XX
PT Monitoring differential expression of genes in filamentous fungal cells
XX uses fluorescence-labeled nucleic acids isolated from the cells and a
XX substrate of expressed sequence tags -

PS
XX
CC The present invention describes a method for monitoring differential
CC expression of genes in a first filamentous fungal (FF) cell relative to
CC expression of the same genes in one or more second filamentous fungal
CC cells. The method uses fluorescence-labeled nucleic acids isolated from
CC the FF cells and a substrate of expressed sequence tags (EST). The ESTs
CC are used in the methods for monitoring differential expression of genes
CC in a first filamentous fungal (FF) cell relative to expression of the
CC same genes in one or more second filamentous fungal cells. Monitoring
CC the global expression of genes from FF cells allows the production
CC potential of the microorganisms to be improved. New genes may be
CC discovered, possible functions of unknown open reading frames can be
CC identified, and gene copy number variation and stability can be
CC monitored. The expression of genes can be used to study how FF cells
CC adapt to changes in culture conditions, environmental stress, spore
CC morphogenesis, recombination, metabolic or catabolic pathway
CC engineering. Using ESTs provides several advantages over genomic or
CC random cDNA clones including elimination of redundancy as one spot on an
CC array equals one gene or open reading frame, and organization of the
CC microarrays based on function of the gene products to facilitate
CC analysis of the results. AAF07478 to AAF11247 represents ESTs from
CC Fusarium venenatum; AAF11248 to AAF11853 represents ESTs from Aspergillus
CC niger; AAF11854 to AAF14878 represents ESTs from Trichoderma reesei; and
CC AAF14879 to AAF15337 represents ESTs from Trichoderma reesei, which are
CC all specifically claimed in the present invention.
XX
SQ Sequence 621 BP; 152 A; 199 C; 123 G; 142 T; 5 other;

Query Match 1.5%; Score 62.8; DB 21; Length 621;
Best Local Similarity 49.4%; Pred. No. 3.2e-06;
Matches 213; Conservative 0; Mismatches 212; Indels 6; Gaps 2;
Qy 576 GTCCAGGCTCCCGCTATGACATCGAGCGGATCAAGCTGCGATTCGTTGCGGTGTA 635
197 GTCCAGGCTCCCGCTATGACATCGAGCGGATCAAGCTGCGATTCGTTGCGGTGTA 256
636 GGTAACCTTTTCGAGCCCATGAGGCGTTCTACGTGCGAGCAGATTCCTTGAACAGCTCCG 695
257 GCGGTTTCCACAGAGCTCACTTGGCTGCTACGTCGACAGAGCTGCTTGAGAAGCAGCGT 316
696 GACTGGCGGATTTGTTGGTGGCTGACGGCGAGTGACCGTTCAAGAAAAAGCCGAG 755
317 GAGCGCGATTTGGGCGCATCTGCGGTATCGGCTGCGTCCCAACGATGTGCCATCGCGAT 376
756 GAATTCAGGCGCCAGGAGCTGCTGTTATTCCTGACCGAGCGGCTCCGTCGCGCAAGAGC 815
377 GTTCTCAACGCGACAGGACCACTCTACACCGTATTTGAGCGATCCCNAAAGGTAGCTTC 436
816 ACGGTGCGCGTCAATGGGCGCGCTGCGTGACTATCTGTTGCCCGCGCGATCCCGAAGCG 875
437 GCGGATGTCGTC---GGCAGCATCACTCTTCTTTCGCGCTGATGACCGTGAGGCT 492
876 GTGCTGAAGCATCTTGTGTGATCCGCGCATCCGATCGTTTCCATGACGATCAGGAAGGC 935
493 GGCATTGCAAGAGTGGNTCATCCGACACTACATTTGTTTTT--TACTATTACTGAGAGN 550
936 GGCTACAACATCAACGAGACGCGGTGCGTTGCGATCTGGAGATTCGCGCAGTAAAGGCC 995
551 GGNTACTACTACACGAGAACACCCATCAATTTGAAGCTGAGCATTTCTGGCATTCAGCAC 610
996 GACCTCAAGAA 1006
611 GACTTCAAGAA 621

RESULT 14

AAS92759/c
ID AAS92759 standard; cDNA: 921 BP.
XX
AC AAS92759;
XX
DT 13-FEB-2002 (first entry)

XX	DNA encoding novel human diagnostic protein #28563.	
DE	Human; chromosome mapping; gene mapping; gene therapy; forensic;	
XX	food supplement; medical imaging; diagnostic; genetic disorder; ss.	
KW	Homo sapiens.	
XX		
XX	WO200175067-A2.	
PN		
XX	11-OCT-2001.	
PD		
XX		
PF	30-MAR-2001; 2001WO-US08631.	
XX		
PR	31-MAR-2000; 2000US-0540217.	
PR	23-AUG-2000; 2000US-0649167.	
XX		
PA	(HYSE-) HYSEQ INC.	
PI	Drmanac RT, Liu C, Tang YT;	
XX		
DR	WPI; 2001-639362/73.	
DR	P-PSDB; ABG28572.	
XX		
PT	New isolated polynucleotide and encoded polypeptides, useful in	
PT	diagnostics, forensics, gene mapping, identification of mutations	
PT	responsible for genetic disorders or other traits and to assess	
PT	biodiversity	
XX		
PS	Claim 1; SEQ ID No 28563; 103pp; English.	
XX		
CC	The invention relates to isolated polynucleotide (I) and	
CC	polypeptide (II) sequences. (I) is useful as hybridisation probes,	
CC	polymerase chain reaction (PCR) primers, oligomers, and for chromosome	
CC	and gene mapping, and in recombinant production of (II). The	
CC	polynucleotides are also used in diagnostics as expressed sequence tags	
CC	for identifying expressed genes. (I) is useful in gene therapy techniques	
CC	to restore normal activity of (II) or to treat disease states involving	
CC	(II). (II) is useful for generating antibodies against it, detecting or	
CC	quantitating a polypeptide in tissue, as molecular weight markers and as	
CC	a food supplement. (II) and its binding partners are useful in medical	
CC	imaging of sites expressing (II). (I) and (II) are useful for treating	
CC	disorders involving aberrant protein expression or biological activity.	
CC	The polypeptide and polynucleotide sequences have applications in	
CC	diagnostics, forensics, gene mapping, identification of mutations	
CC	responsible for genetic disorders or other traits to assess biodiversity	
CC	and to produce other types of data and products dependent on DNA and	
CC	amino acid sequences. AAS64197-AAS94564 represent novel human	
CC	diagnostic coding sequences of the invention.	
CC	Note: The sequence data for this patent did not appear in the printed	
CC	specification, but was obtained in electronic format directly from WIPO	
CC	at ftp.wipo.int/pub/published_pct_sequences .	
XX		
XX	Sequence 921 BP; 230 A; 235 C; 230 G; 226 T; 0 other;	
QY	Query Match 1.3%; Score 53.8; DB 23; Length 921;	
Db	Best Local Similarity 53.6%; Pred. No. 0.0009;	
QY	Matches 112; Conservative 0; Mismatches 97; Indels 0; Gaps 0;	
Db	1307 CGAGGATTTCCATCAGTCGGGTGCTGGAAGACCACTTTGCGGATGCGCGCTCGGATGGA 1366	
QY	1307 CGAGGATTTCCATCAGTCGGGTGCTGGAAGACCACTTTGCGGATGCGCGCTCGGATGGA 848	
Db	1367 AAAAGCCGCGCTGCAGATGCTGCGGGACGTCGAGGACTGGGAGTACGTCAGATCCGAT 1426	
QY	847 AGTCGCAAGTGTACAAATGGTGAATGATGCTTCGCCATGGGAAGATGAAGATCGCGAT 788	
Db	1427 GCTCAATGACGGGATGTGTCATGCTCTGCTTCCAGGCAATTCCTGGTCGGCTATGAGAATCT 1480	
QY	787 GCTTAATGGCAGCCACTCTTTTCTCGCTTATCTGGGTACCTCTCAGGATTCGCCATAT 728	
Db	1487 GGATGACGCCATTTGAAGACAGCAGCAACTCC 1515	

```
QY 1382 GATGTCGGGACGTGACGGACTGGGAGTAGCTCAAGATCGAATGCTCAATGCAGGGCA 1441
      |||||  |||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
Db 379 AATGGTGAATGATGCTCCTGCCATGGGAGAGATGAACTGCGGATGCTTATATGCAGCCA 320
      |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
QY 1442 TGTCAATGCTCTGCTCCAGGCATTCGGTCGGCTATGAGAAATGGATGACGCCATTGA 1501
      |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
Db 319 CTCCTTTCTCGCTTATCTGGGTTACCTCTCAGGATTCGCCCATATCAGTGATTGCATGCA 260
      |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
QY 1502 AGACAGCGAATCCTTGGCAATCTGAAGAATATCTCAACAAGGATGTCATCCCGACCCCT 1561
      |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
Db 259 GGATCGCCGCA---TTTCGCCATGCCGCCAGAACATTAATGCTGGATGAGCAAGCGCCGAC 203
      |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
QY 1562 GAAGGCGCCTTCAGGCATGACGCTCGAAGGCTATCGGGACAGCGTCAATCAGCCGTTTCTC 1621
      |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
Db 202 ACTGCAAAATTAAGATGTCGATTTAACACAAATATCCGGATTAAGTTAATTGCACGTTTGC 143
      |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
QY 1622 CAACAAGCGGATGTCGGACGACAGCGCTCCGGATTGCTAGCGATGGCTGTTCCAAAGGTTCA 1681
      |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
Db 142 TAATCCGGCGCTGAAACATAAGACCTGGCAAAATCCGATGGATGGCAGCCCAAGAAATTACC 83
      |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
QY 1682 GGTGTTCTGGACGGAACCGTGGCGGCGATCGAAGACAAAGCGGACCTGTTCACGTAT 1741
      |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
Db 82 GCACCGCATGCTGGCAGGTATTCGCATACATCAAGGGCGCCGAACGGACTGTCGTTGCT 23
      |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
QY 1742 AGCGTTCGGAATTCATCCTA 1762
      |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
Db 22 GGCATTACGAGTCACGTCTTA 2
      |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
```

Search completed: May 18, 2003, 23:47:14
Job time : 1197 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: May 18, 2003, 16:35:29 ; Search time 177 Seconds
(without alignments)
7129.802 Million cell updates/sec

Title: US-09-926-163B-1
Perfect score: 4115
Sequence: 1 aagcttcagctgcctgcaggt.....tccgggcccgcctgaagctt 4115

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 441362 seqs, 153338381 residues

Total number of hits satisfying chosen parameters: 882724

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : Issued_Patents_NA.*

- 1: /cgn2_6/ptodata/1/ina/5A_COMB.seq.*
- 2: /cgn2_6/ptodata/1/ina/5B_COMB.seq.*
- 3: /cgn2_6/ptodata/1/ina/6A_COMB.seq.*
- 4: /cgn2_6/ptodata/1/ina/6B_COMB.seq.*
- 5: /cgn2_6/ptodata/1/ina/PCTUS_COMB.seq.*
- 6: /cgn2_6/ptodata/1/ina/backfiles.seq.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
C 1	44	1.1	1926	4	US-09-249-585A-4
C 2	44	1.1	1931	2	US-09-130-114-2
C 3	43	1.0	477	4	US-09-135-994-1
C 4	42.6	1.0	2159	3	US-08-286-870A-7
C 5	41.2	1.0	533	6	5482709-5
C 6	40.4	1.0	543	6	5273901-6
C 7	40	1.0	234	1	US-08-469-802B-3
C 8	40	1.0	234	2	US-08-267-803B-3
C 9	39.6	1.0	4403765	4	US-09-103-840A-2
C 10	39.6	1.0	4411529	4	US-09-103-840A-1
C 11	39.4	1.0	298	2	US-08-637-759B-21
C 12	39.4	1.0	298	3	US-08-871-355A-21
C 13	39.4	1.0	298	4	US-09-201-945-21
C 14	38.8	0.9	397	3	US-09-232-691-3
C 15	38.6	0.9	1794	4	US-09-232-191-36
C 16	38.6	0.9	1794	4	US-09-232-200-90
C 17	38.6	0.9	1794	4	US-09-232-197-90
C 18	38.6	0.9	1794	4	US-09-232-201-90
C 19	38.6	0.9	2007	4	US-09-232-191-22
C 20	38.6	0.9	2007	4	US-09-232-200-22
C 21	38.6	0.9	2007	4	US-09-232-197-22
C 22	38.6	0.9	2007	4	US-09-232-197-22
C 23	38.6	0.9	8540	4	US-08-487-283A-4
C 24	38.6	0.9	8540	5	PCT-US96-05611A-12
C 25	38.6	0.9	8932	2	US-08-252-493C-8
C 26	38.6	0.9	8932	3	US-09-276-197-8
C 27	38.6	0.9	4403765	4	US-09-103-840A-2

28	38.6	0.9	4411529	4	US-09-103-840A-1	Sequence 1, Appli
29	38.4	0.9	2870	5	PCT-US93-07213-3	Sequence 3, Appli
30	38.2	0.9	2799	4	US-09-232-279-1	Sequence 1, Appli
31	38.2	0.9	3382	2	US-08-682-847-1	Sequence 1, Appli
C 32	38	0.9	3010	4	US-09-199-637A-168	Sequence 168, App
33	37.8	0.9	279	4	US-09-481-288-14	Sequence 14, Appli
34	37.8	0.9	932	1	US-08-114-692A-3	Sequence 3, Appli
35	37.8	0.9	932	2	US-08-723-306-3	Sequence 3, Appli
36	37.8	0.9	932	5	PCT-US96-10041-3	Sequence 3, Appli
C 37	37.8	0.9	1693	3	US-09-320-878-23	Sequence 23, Appli
38	37.8	0.9	30001	1	US-08-125-468-1	Sequence 1, Appli
39	37.8	0.9	30001	2	US-08-474-933-1	Sequence 1, Appli
C 40	37.6	0.9	3366	1	US-08-469-802B-1	Sequence 1, Appli
C 41	37.6	0.9	3366	2	US-08-267-803B-1	Sequence 1, Appli
C 42	37.6	0.9	10348	2	US-08-457-273B-41	Sequence 41, Appli
C 43	37.6	0.9	10348	3	US-08-556-419-13	Sequence 13, Appli
C 44	37.6	0.9	10348	4	US-09-041-886-14	Sequence 14, Appli
C 45	37.6	0.9	10366	1	US-08-246-982A-5	Sequence 5, Appli

ALIGNMENTS

RESULT 1
US-09-249-585A-4/c
; Sequence 4, Application US/09249585A
; Patent No. 6417002
; GENERAL INFORMATION:
; APPLICANT: Horlick, Robert
; TITLE OF INVENTION: METHOD FOR MAINTENANCE AND SELECTION OF EPISOMES
; FILE REFERENCE: 0867/0D905
; CURRENT APPLICATION NUMBER: US/09/249,585A
; CURRENT FILING DATE: 1999-02-11
; NUMBER OF SEQ ID NOS: 18
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 4
; LENGTH: 1926
; TYPE: DNA
; ORGANISM: Epstein Barr Virus
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (1)..(1926)
; OTHER INFORMATION: template strand of EBNA-1 DNA
US-09-249-585A-4

Query Match	1.18;	Score 44;	DB 4;	Length 1926;
Best Local Similarity	48.18;	Pred. No. 0.031;		
Matches 125;	Conservative 0;	Mismatches 135;	Indels 0;	Gaps 0;
Qy	1154	GAAGCGCGGATCCGGAGTTGGCGAAGTGTGAGGAAACCGGACCTTCCCGAACGG	1213	
Db	1022	GGAGGAGCTGGGCGCGGAGGTGACGGAGGAGTGGGCGCGGAGTGGAGGACGAGGACGG	963	
Qy	1214	AATGTTGATGCGATACCCCGGACCGTTTCGGCGGAATCCCAAGAGCTCAACGCGGC	1273	
Db	962	GGAGGACGACGACCGGGGAGGAGGAGGACGGGAGGACGGGAGGACGAGGAGGAGGA	903	
Qy	1274	CAGTGGCTGGATCAGCACCTCGCTGGTGGCGGAGGATTTCCATCAGTGGGTCTGGA	1333	
Db	902	GGACGGGAGGACGAGGACGGGAGGAGGAGGACGGGAGGACGAGGAGGAGGAGGA	843	
Qy	1334	AGACGAGTTTGGGATGCGCGCTTCGCGCTTTGAAAAACCGCGCTGCAGATGTTCTCGGGA	1393	
Db	842	GGACGAGGACGGGAGGACGGGAGGAGGAGGACGAGGACGGGAGGACGAGGAGGA	783	
Qy	1394	CGTGACGACTGGGAGTACG	1413	
Db	782	CGGGAGGACGGGAGGACG	763	

RESULT 2
US-09-130-114-2/c
; Sequence 2, Application US/09130114

```
; Patent No. 5976807
; GENERAL INFORMATION:
; APPLICANT: Horlick, Robert A.
; APPLICANT: Dama, Bassam B.
; APPLICANT: Robbins, Alan K.
; TITLE OF INVENTION: Eukaryotic Cells Stably Expressing Genes
; FILE REFERENCE: 0867/ID903US1
; CURRENT APPLICATION NUMBER: US/09/130,114
; CURRENT FILING DATE: 1998-08-06
; NUMBER OF SEQ ID NOS: 36
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 2
; LENGTH: 1931
; TYPE: DNA
; ORGANISM: EBNA
US-09-130-114-2

Query Match      1.1%; Score 44; DB 2; Length 1931;
Best Local Similarity 48.1%; Pred. No. 0.031;
Matches 125; Conservative 0; Mismatches 135; Indels 0; Gaps 0;

QY 1154 GAAGCGCGCCATCCGAGTTGGCGAAGTGTGAGGAAACCGACCTTCCCGACGG 1213
      ||||| || ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 1022 GGAGGAGCTGGGGCGGAGGTGACGAGAGTGGCGGCGGAGGTGGAGGACGAGCGG 963
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 1214 AATGTTGATCGCATCACCCGACCGTTCGCGCGAAATCCCAAGAGTCAACGCGGC 1273
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 962 GGAGGAGGACGCGGGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGGA 903
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 1274 CAGTGGGCTGATGACGACCTGCGCTGGTGGCGCGAGGATTTCCATCAGTGGGTGGA 1333
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 902 GGACGGGAGGACGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGA 843
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 1334 AGACGAGTTCCGATGCGCTCGCGCTTGAAGAACCGCGCTGCAGATGGTTCGGGA 1393
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 842 GGACGAGGACGCGGGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGGA 783
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 1394 CGTGACGAGCTGGGAGTACG 1413
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 782 CGGGGAGGACGCGGGAGGACG 763
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

RESULT 3
US-09-135-994-1/c
; Sequence 1, Application US/09135994A
; Patent No. 6280938
; GENERAL INFORMATION:
; APPLICANT: Rasmus et al.
; TITLE OF INVENTION: SCA7 GENE AND METHODS OF USE
; FILE REFERENCE: University of Minnesota
; CURRENT APPLICATION NUMBER: US/09/135,994A
; CURRENT FILING DATE: 1998-08-18
; EARLIER APPLICATION NUMBER: 60/056,170
; EARLIER FILING DATE: 1997-08-19
; NUMBER OF SEQ ID NOS: 14
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 1
; LENGTH: 477
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-135-994-1

Query Match      1.0%; Score 43; DB 4; Length 477;
Best Local Similarity 47.9%; Pred. No. 0.028;
Matches 124; Conservative 0; Mismatches 135; Indels 0; Gaps 0;

QY 3357 CGCTGTCGAATGGCTGCGAGAAGTGTGAACTGGTTGGCGGAAGCGGATCAATAATG 3416
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 316 GCTGTCGCGCTGGGGGTGCGGAGGCGCGGCTGCTGCTGCTGCTGCTGCTGCTGCT 257
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 3417 GGTTCGATTCGCGGCGGCGTGTTCATCTGCGCGCGGCTTCTGTCATCTGCGCTGT 3476
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
```

```
DB 256 GCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 197
QY 3477 TCTGGGTCTGTGGATGCTGCGGGTCCGGTGTGCTGCTGCTGCTGCTGCTGCTGCTT 3536
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 196 GCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 137
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 3537 CCCTGCGCTTCAGGGCGCGCTCTTTCGCTGGATCGAAGTCAACATCCGATTCGGTGG 3596
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 136 GCTGCTGCTGCGGGCGCGCTGCTCCGCGCGCGCGCGCGCGCGCGCGCGCGCTCCC 77
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 3597 CCTTCCGGCGAAGAGCGGC 3615
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 76 CCCTGACGTATCCTCGCGGC 58
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

RESULT 4
US-08-286-870A-7/c
; Sequence 7, Application US/08286870A
; Patent No. 6063605
; GENERAL INFORMATION:
; APPLICANT: ELY, S
; APPLICANT: TAILOR, RH
; APPLICANT: TIPPETT, JM
; APPLICANT: BLENN, RG
; TITLE OF INVENTION: BACTERIAL GENES
; NUMBER OF SEQUENCES: 10
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: CUSHMAN DABY & CUSHMAN
; ADDRESSEE: Intellectual Property Group of
; STREET: 1100 North Washington, N.W.
; CITY: Washington
; STATE: DC
; COUNTRY: USA
; ZIP: 20005-3918
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; OPERATING SYSTEM: IBM PC compatible
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/286,870A
; FILING DATE: 05-AUG-1994
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/520228
; FILING DATE: 09-MAY-1990
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: GB 8910624.9
; FILING DATE: 09-MAY-1989
; ATTORNEY/AGENT INFORMATION:
; NAME: PAUL N. KOKULIS
; REGISTRATION NUMBER: 16,773
; REFERENCE/DOCKET NUMBER: 70608/220720
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (202) 861-3000
; TELEFAX: (202) 822-0944
; INFORMATION FOR SEQ ID NO: 7:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 2159 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; MOLECULE TYPE: cDNA
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 1..2159
US-08-286-870A-7

Query Match      1.0%; Score 42.6; DB 3; Length 2159;
Best Local Similarity 54.0%; Pred. No. 0.081;
Matches 87; Conservative 0; Mismatches 74; Indels 0; Gaps 0;
```

2699	QY	ACATCTCGAANTCGTCCGTCGGTCCGGGCGCGGATCGAACAGATGCCACATTCCTTGG	2758
712	Db	ACTTCAGCAANTGGTCGCTAGTCGCGCGCGCTCCACCTGCCTGTTGTAGAACGTGC	653
2759	QY	TGTATTGGGGGGAACTGGAAACAGCTCTTAAAGCGTTGATTTCGTGTCGGTTCACCG	2818
652	Db	TGATCTCGTGTGGAGACAGGCCCCACTCTTTCGCGAAGATGCTGGGCTCGCAGAGACA	593

RESULT 5

RESOL 3
5482709-5/c
; Patent NO. 5482709
; APPLICANT: JACOBSON, JAMES W.; STRAUSBERG, ROBERT L.; WILSON,
; SUSAN D.; POPE, SHARON H.; STRAUSBERG, SUSAN L.; RUFF, MICHAEL D.;
; AUGUSTINE, PATRICIA C.; DANFORTH, HARRY D.
; TITLE OF INVENTION: EIMERIA ANTIGENIC COMPOSITION WHICH
; ELICITS ANTIBODIES AGAINST AVIAN COCCIDIOSIS
; NUMBER OF SEQUENCES: 10
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/148,432
; FILING DATE: 08-NOV-1993
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 581,693
; FILING DATE: 12-SEP-1990
; APPLICATION NUMBER: 215,162
; FILING DATE: 05-JUL-1989
; APPLICATION NUMBER: 746,520
; FILING DATE: 19-JUN-1985
; APPLICATION NUMBER: 627,811
; FILING DATE: 05-JUL-1984
; SEQ ID NO: 5
; LENGTH: 533
5482709-5

```

: FILING DATE: 05-JUL-1988
: APPLICATION NUMBER: 748 520
: FILING DATE: 19-JUN-1985
: APPLICATION NUMBER: 627 811
: FILING DATE: 05-JUL-1984
: SEQ ID NO: 6
: LENGTH: 543
5273901-6

```

9 JUL 68

RESULT 6
 5273901-6/c
 ; Patent No. 5273901
 ; APPLICANT: JACOBSON, JAMES W.; STRAUSBERG, ROBERT L.; WILSON,
 ; SUSAN D.; POPE, SHARON H.; STRAUSBERG, SUSAN L.; RUFF, MICHAEL D.;
 ; AUGUSTINE, PATRICIA C.; DANFORTH, HARRY D.
 ; TITLE OF INVENTION: GENETICALLY ENGINEERED COCCIDIOSIS
 ; SPOROZOITE 21.5 KB ANTIGEN, AC-6B
 ; NUMBER OF SEQUENCES: 11
 ; CURRENT APPLICATION DATA:
 ; APPLICATION NUMBER: US/07/581,693
 ; FILING DATE: 12-SEP-1990
 ; PRIOR APPLICATION DATA:
 ; APPLICATION NUMBER: 215,162


```

RESULT 12
US-08-871-355A-21
; Sequence 21, Application US/08871355A
; Patent No. 6015669
; GENERAL INFORMATION:
; APPLICANT: David William Holden
; TITLE OF INVENTION: Identification of Genes
; NUMBER OF SEQUENCES: 501
; CORRESPONDENCE ADDRESS:

```

ADDRESSEE: Patrea L. Pabst
STREET: 2800 One Atlantic Center
STREET: 1201 West Peachtree Street
CITY: Atlanta
STATE: Georgia
COUNTRY: USA
ZIP: 30309-3450
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/871.355A
FILING DATE: 09-JUN-1997
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: PCT/GB95/02875
FILING DATE: 11-DEC-1995
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: Pabst, Patrea L.
REGISTRATION NUMBER: 31,284
REFERENCE/DOCKET NUMBER: RPMS 101 CON
TELECOMMUNICATION INFORMATION:
TELEPHONE: (404) 873-8794
TELEFAX: (404) 873-8795
INFORMATION FOR SEQ ID NO: 21:
SEQUENCE CHARACTERISTICS:
LENGTH: 298 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
HYPOTHETICAL: NO
ANTI-SENSE: NO
ORIGINAL SOURCE:
ORGANISM: Partial sequence of Salmonella typhimurium
ORGANISM: virulence gene
US-08-871-355A-21

	Query Match	1.08;	Score 39.4;	DB 3;	Length 298;
	Best Local Similarity	59.3%;	Pred. No. 0.22;		
	Matches 67;	Conservative 0;	Mismatches 46;	Indels 0;	Gaps 0;
QY	1	AAGCTTTGCATCGCCTCGAGGTCGACTCTAGAGGATCCGGTTTTTGSCAGCGCTCCCTAGATT	60		
Db	12	AAGCTTTGCATCGCCTCGAGGTCGACTCTAGAGGATCTACTAGTCATATGGATTTCTCTAGCGC	71		

RESULT 14
US-09-253-691-3/c
; Sequence 3, Application US/09253691
; Patent No. 6124100

	Query Match	Best Local Similarity	Score 38.6;	DB 4;	Length 1794;
	Matches 95;	Conservative 0;	Mismatches 0;	Indels 94;	Gaps 0;
QY	1139	CTTCTTCGGGTATGCGAAGCGCGCATCCGGAGTTGGCGAAGTGGATTGAGAAACGC	1198		
Db	1578	CTTGC CGCGGTATGCGACTCTTTGTTCTGGGTAGTGGGTCTGTCGGCGACACCAC	1537		
QY	1199	GACCTTTCCCGAACCGGAATGGTTGATCGATCAACCCCGACCGTTTCGCGCGAAATCGCAA	1258		

Db	1638	GAAGCTTCAAGAGTCGCAAGGTGGAGTTGGCCACACAGGCCTATGGCGCCGACATCGAGGA	1697
Qy	1259	GAAGCTCAACCGCCGACATGGCTGGATGACGACCTGCCCGCTGGTGGCGGAGGATTCCA	1318
Db	1698	TCCGCTGTAGCTACTGGCCGCCGACGAGGATATGCCCGTACTACGCCGATACCC	1757
Qy	1319	TCAGTGGGT	1327
Db	1758	TGAGGAGGT	1766

Search completed: May 18, 2003, 22:35:23
Job time : 7951 secs

THIS PAGE BLANK (USPTO)

GenCore version 5.1.6.
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: May 18, 2003, 17:05:15 ; Search time 487 Seconds
(without alignments)
10899.580 Million cell updates/sec

Title: US-09-926-163B-1
Perfect score: 4115
Sequence: 1 aagcttgatcgctgcaggt.....tccggcgccgctgaagctt 4115

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 810007 seqs, 644969091 residues

Total number of hits satisfying chosen parameters: 1620014

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : PublishedApplications_NA:*

1: /cgn2_6/ptodata/1/pubpna/US07_PUBCOMB.seq:*
2: /cgn2_6/ptodata/1/pubpna/US07_PUBCOMB.seq:*
3: /cgn2_6/ptodata/1/pubpna/US06_NEW_PUB.seq:*
4: /cgn2_6/ptodata/1/pubpna/US06_PUBCOMB.seq:*
5: /cgn2_6/ptodata/1/pubpna/US07_NEW_PUB.seq:*
6: /cgn2_6/ptodata/1/pubpna/US07_PUBCOMB.seq:*
7: /cgn2_6/ptodata/1/pubpna/US08_NEW_PUB.seq:*
8: /cgn2_6/ptodata/1/pubpna/US08_PUBCOMB.seq:*
9: /cgn2_6/ptodata/1/pubpna/US09_NEW_PUB.seq:*
10: /cgn2_6/ptodata/1/pubpna/US09_PUBCOMB.seq:*
11: /cgn2_6/ptodata/1/pubpna/US10_NEW_PUB.seq:*
12: /cgn2_6/ptodata/1/pubpna/US10_PUBCOMB.seq:*
13: /cgn2_6/ptodata/1/pubpna/US60_NEW_PUB.seq:*
14: /cgn2_6/ptodata/1/pubpna/US60_PUBCOMB.seq:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match %	Length	DB ID	Description
1	213.6	5.2	1509	9	US-09-738-626-126
2	213.6	5.2	3309400	9	US-09-738-626-1
3	105	2.6	1848	9	US-09-802-208B-1
4	42	1.0	2371	10	US-09-789-561-27
5	41.6	1.0	2614	9	US-09-822-846-491
6	41.4	1.0	1512	10	US-09-822-849A-393
7	41.4	1.0	1575	10	US-09-822-849A-261
8	41	1.0	501	10	US-09-833-790-158
9	40.4	1.0	1432	10	US-09-974-300-718
10	40.4	1.0	8017	9	US-10-152-040-26
11	40.4	1.0	8092	9	US-10-152-040-27
12	40.4	1.0	10417	9	US-10-152-040-28
13	40.4	1.0	536165	9	US-09-939-964-1
14	40.2	1.0	88421	9	US-09-976-059-1
15	40	1.0	3115	9	US-09-981-353-85
16	40	1.0	6604	10	US-09-880-107-1748
17	39.6	1.0	1146	9	US-09-738-626-2512
18	39.6	1.0	1854	9	US-09-894-844-103
19	39.6	1.0	3309400	9	US-09-738-626-1

20	38.8	0.9	724	9	US-10-123-155-60
21	38.6	0.9	1794	10	US-09-943-671-36
22	38.6	0.9	2007	10	US-09-943-671-22
23	38.4	0.9	2283	9	US-09-712-363-135
24	38.2	0.9	440	9	US-10-184-644-202
25	38.2	0.9	440	9	US-10-184-634-202
26	38.2	0.9	2799	9	US-10-085-519-1
27	38.2	0.9	5607	9	US-10-108-605-92
c 27	38.2	0.9	5607	9	US-10-123-155-358
c 28	38	0.9	3010	9	US-09-975-719-168
c 29	38	0.9	3010	9	US-09-910-009A-157
30	37.4	0.9	353	9	US-10-184-644-60
31	37.4	0.9	694	9	US-10-184-634-60
32	37.4	0.9	694	9	US-10-184-634-104
33	37.2	0.9	415	10	US-09-960-352-1970
34	37.2	0.9	507	10	US-09-894-882-8
35	37.2	0.9	1503841	9	US-09-946-807-1
36	37.2	0.9	1503841	10	US-09-795-668-1
37	37.2	0.9	1503841	10	US-09-795-668-1
c 38	37	0.9	750	9	US-10-184-644-104
c 39	37	0.9	750	9	US-10-184-634-104
c 40	37	0.9	3796	9	US-10-127-391-32
c 41	36.8	0.9	18272	9	US-09-764-891-8869
c 42	36.6	0.9	941	9	US-10-123-155-464
c 43	36.6	0.9	3653	9	US-10-155-649-1
c 44	36.6	0.9	13910	9	US-09-919-901-1
c 45	36.6	0.9	13910	9	US-09-919-901-8

ALIGNMENTS

RESULT 1

US-09-738-626-126
Sequence 126, Application US/09738626
Publication No. US20020197605A1

GENERAL INFORMATION:
APPLICANT: NAKAGAWA, SATOSHI

APPLICANT: MIZOGUCHI, HIROSHI

APPLICANT: ANDO, SEIKO

APPLICANT: HAYASHI, MIKIRO

APPLICANT: OCHIALI, KEIKO

APPLICANT: YOKOI, HARUHIKO

APPLICANT: TATEISHI, NAKO

APPLICANT: SENO, AKIHIRO

APPLICANT: IKEDA, MASATO

APPLICANT: OZAKI, AKIO

TITLE OF INVENTION: NOVEL POLYNUCLEOTIDES

FILE REFERENCE: 249-125

CURRENT APPLICATION NUMBER: US/09/738,626

CURRENT FILING DATE: 2000-12-18

PRIOR APPLICATION NUMBER: JP 99/377484

PRIOR FILING DATE: 1999-12-16

PRIOR APPLICATION NUMBER: JP 00/159162

PRIOR FILING DATE: 2000-04-07

PRIOR APPLICATION NUMBER: JP 00/280988

PRIOR FILING DATE: 2000-08-03

NUMBER OF SEQ ID NOS: 7059

SOFTWARE: PatentIn ver. 3.0

SEQ ID NO 126

LENGTH: 1509

TYPE: DNA

ORGANISM: Corynebacterium glutamicum

US-09-738-626-126

Query Match 5.2%; Score 213.6; DB 9; Length 1509;

Best Local Similarity 49.6%; Pred. No. 4.2e-56;

Matches 617; Conservative 0; Mismatches 609; Indels 18; Gaps 2;

QY 568 CTGCCAATGTCAGGCTCCCTTATGACATCGACGGGATCAAGCCTGGGATCGTGATT 627

Db 53 CTTCGGGAGTGCAGATCCCGGCTTCAACCGCGCTGACGTGCGCCGGGATGTCCTACT 112

QY 628 TCGGTGTAGTAACTTTTTCGAGCCCATGAGGCGTTCTACGTGAGCAGATTCTTGAAC 687

Db 113 TCGGTGTGGGGATTCATCGCGCTACCAAGGATGTACCTCAATGAATTGATG 172
Qy 688 ACGTCCGGAGTGGCGATTGTTGGTGTGGCTGACGGGAGTGACCGTTCAAGAAA 747
Db 173 AGGCAAGGCGCTTGATTGGGCGCATATCGGCGATGGGTGTCATGCCCTCCGATGCGCA 232
Qy 748 AAGCGAGGAATCAAGGCCAGGACTGCCGTGATTCCTCAGCAGGAGCGCTCCGTCG 807
Db 233 TCGCGATGCCCTGGCCAGCAAGATCACCTTTATACCTGACCACTAAAGCTCCTGATG 292
Qy 808 GCAAGAGACAGGTGGCGGTATGGCGCGCTGCGTACTATCTCTTTCGCCCGCCGATC 867
Db 293 GAATCTTTGATCAAAAAATATCGATCCATCATTTAGTACGTGCTGCTCCCGAGGAC 352
Qy 868 CGGAAGCGGTGCTGAAGCATCTTTGATCGCGGCATCCGATCGTTTCCATGACGATCA 927
Db 353 CAGCAGGGCGGTTGCAACCCCTCGCGAGGACTCCATCCGATTTGCTCCCTCAGGTGA 412
Qy 928 CGGAAGCGGTGCTGAAGCATCTTTGATCGCGGCATCCGATCGTTTCCATGACGATCA 987
Db 413 CTGAAGCGGTGCTGAAGCATCTTTGATCGCGGCATCCGATCGTTTCCATGACGATCA 472
Qy 988 TAAAGCGGACTCAAGAACCGGAAAGCGGTCT-----ACCGTTTTCG 1032
Db 473 TCGTTGCTGACCGGAGCGCTGACGGCGGCGATATCTCCACTTTGCGAGACCTTCTTG 532
Qy 1033 GTTACGTGGTGGAGCGCTGCGTCTGTTGGGATGCGGTTGGTAAGGCAATTTACGGTCA 1092
Db 533 GTTATGATCAGTGGCGCATTTGTTCCCGAAAGAAATCAGGATCTACGCCATTTACCATCA 592
Qy 1093 TGTCTGTGATAACCTGCTCATACGCAATGTCGCCGAGCGCTTCTCCGCGTATG 1152
Db 593 TGAGCTGCGATACATCAAGCAACGCGATCTGCTAAGGTTTCTTCTCGCTCG 652
Qy 1153 CAAAGCGCGGATCGCGGATGTCGGAAGTGGATGAGGAAAGCGACCTTTCCCGAAG 1212
Db 653 CACATTCGCTGTCTCTGAGTGGCGAATGGTGGAAACAAAGTGGCTTCCCAACT 712
Qy 1213 GAATGTTGATGCGATACCGCGCTTTCGGCGGAATGCCAAGAGCTCAACGGG 1272
Db 713 CCATGTGGAGCGCATACCC---CTGAACACCGAGCGGCGGCGGATGACATCAAG 769
Qy 1273 CCAGTGGGCTGGATGACGACCTGCGCTGTGGCGGAGGATTTCCATCAGTGGGTGCTG 1332
Db 770 AAATCGGCTACATCGATCGTGGCGAGTGTCTGGAAGATTTCAACCAATGGGTCTCG 829
Qy 1333 AAGACAGTTTGGGATGCGGCTGCGCGCTTGAAGAAAGCGGCGTGCAGATGTCGGG 1392
Db 830 AGGATGCTTTCACCGAGGCGCGCGCTACGAGAGGTTGGCGTGCAGTCTCTCCG 889
Qy 1393 ACGTGACGAGCTGGGAGTACGTCAAGATCCGAATGCTCAATGCGAGGCGATGTCATGCT 1452
Db 890 ACGTGAGCGCTTATGAATTAATGAAGCTGCGCTGCTCAAGCGCTCCACCGAGGACTT 949
Qy 1453 GCTTCCCGAGGATCTGCTGCGCTATGAGATGTTGGATGACGCCATTAAGACAGCGAAC 1512
Db 950 GCTACTTCCGGCACTTGGCTGGCCACACATGTTCCAGAGCTCATGCGGATACCCGCT 1009
Qy 1513 TCTTGGCACTGAGAGACTATCTACACAGGATGTCATCCGACCGCTGAAGCGGCTT 1572
Db 1010 TCCAGGATTTCTCTGCTTACATGAGCGGAGGAGCGGAGCGGCGGCTTCAAGGAACTTC 1069
Qy 1573 CAGGATGAGCTTCAAGGCTATCGGAGCGGCTCATGCGGCTTCTTCCCAAGAGGGA 1632
Db 1070 CAGGTGTCGATGATGCTTATGACCGCAACTCATCGCGGATTCGGCAACGCGGAG 1129
Qy 1633 TGTCCGACAGAGCTCCCGATGCTAGCGATGGCTTTCGAAGGTTTCAGGTGTTCTGGA 1692
Db 1130 TCAAGACACAGCTACCGCGCTGTGTGCGGAATCTCCGACGCGATTTCCAAAGTGGCTG 1189
Qy 1693 CGGAACAGCTGGTGGCGGATTCGAAGACAGCGGAGCTGTCAGTATACGGTTCCGAA 1752

Db 1190 TGCAGTCTGACGCGAAGAACTTCGACAGAGCCCGGAGCTCACACTTTCTGACGCCATCG 1249
Qy 1753 TTGCATCTCTATCGAAATGCTGGTGTGGCTGCGGAGAGAGGCG 1796
Db 1250 TCGCATCTGGCGCGCTACGCAAGAGCCAGCAGGAGGCG 1293
RESULT 2
US-09-738-626-1/c
; Sequence 1, Application US/09738626
; Publication No. US20020197605A1
; GENERAL INFORMATION:
; APPLICANT: NAKAGAWA, SATOSHI
; APPLICANT: MIZOGUCHI, HIROSHI
; APPLICANT: ANDO, SEIKO
; APPLICANT: HAYASHI, MIKIRO
; APPLICANT: OCHIALI, KEIKO
; APPLICANT: YOKOI, HARUHIKO
; APPLICANT: TATEISHI, NAOKO
; APPLICANT: SENOH, AKIHIRO
; APPLICANT: IKEDA, MASATO
; APPLICANT: OKAZAKI, AKIO
; TITLE OF INVENTION: NOVEL POLYNUCLEOTIDES.
; FILE REFERENCE: 249-125
; CURRENT APPLICATION NUMBER: US/09738,626
; CURRENT FILING DATE: 2000-12-18
; PRIORITY APPLICATION NUMBER: JP 99/377484
; PRIORITY FILING DATE: 1999-12-16
; PRIORITY APPLICATION NUMBER: JP 00/159162
; PRIORITY FILING DATE: 2000-04-07
; PRIORITY APPLICATION NUMBER: JP 00/280988
; PRIORITY FILING DATE: 2000-08-03
; NUMBER OF SEQ ID NOS: 7059
; SOFTWARE: PatentIn ver. 3.0
; SEQ ID NO 1
; LENGTH: 3309400
; TYPE: DNA
; ORGANISM: Corynebacterium glutamicum
US-09-738-626-1

Query Match 5.2%; Score 213.6; DB 9; Length 3309400;
Best Local Similarity 49.6%; Pred. No. 1.1e-54;
Matches 617; Conservative 0; Mismatches 609; Indels 18; Gaps 2;
Qy 568 CTGCGAATTCACAGGCTCCCGCTATGACATCGACGGGATCAAGCTGGGATTCGTGCTT 627
Db 122407 CTTCGGAGTCCAGTCCAGCGTTCACCGCGCTGACGTGCGCCGCGCATTTGCCACT 122348
Qy 628 TCGGTGTAGTAACTTTTTCGAGCCCATGAGGCGTTCAGTCGAGCAGATTTCTGAAC 687
Db 122347 TCGGTGTGGCGGATTCATCGCGCTCACCAAGCGATGTACTCTCAATGAATTTGATGAATG 122288
Qy 688 ACGTCCGAGTGGCGGATTTGTTGGCTTGGCGGAGTACCGTTCAAGAAAAA 747
Db 122287 AGGCAAGCGCTTGGATTGGGCGCATCATCGCATGGGTGTCATGCCCTTCCGATGCGCA 122228
Qy 748 AAGCCGAGGAATCAAGGCCCGAGGACTGCCCTGTTATTCCTCAGCGAGACGCTCCGTCG 807
Db 122227 TCGCGATGCTTGGCGAGCCCAAGATCACCTTTATACCTCAGCCACTAAAGCTCTCTGATG 122168
Qy 808 GCAAGACAGCGTGGCGGCTCATGCGGCGCTGCGTGACTATCTGCTGCCCGGCGGATC 867
Db 122167 GAATCTTTGATCAAAAAATCATCGGATCCATCATTTGACTAGTGTTCCTCCCGAGGACC 122108
Qy 868 CGGAAGCGCTGCTGAAGCATCTTGTGATCGGCGCATCCGCTGCTTTTCCATGACGATCA 927
Db 122107 CAGCAGGCGCTTGCACACCTTCGCGGAGGACTCCATCCGATGTTTCCCTCAGGTTGA 122048
Qy 928 CGGAAGCGCTTACAACTAACACAGACGCGGCTGCTGATCTCGATCTCGAGATCGGCGAG 987
Db 122047 CTGAAGCGGATACAACTCATGATCGGCGGAGAGATTTTCACACACCAACCCCTCGAA 121988
Qy 988 TAAAGCGCGCTCAAGAACCCGGAAGCGCTCT-----ACCGTTTTCG 1032

```

Db      121987 TCGTTGCTGACCGGAGCCCTGACGGGGCGATACCTTCCACTTTGCGAGACCTTCTTTG 121928
QY      1033 GTTACGTGGTCGAGGCGCTGCGTCTGTTGGGTCGCGGTGAAGCATTTACGGTCA 1092
Db      121927 GGTGATCAGTCGCGCATTTGATTTCCGGAAGAAATCAGGATCTACGCCATTTACCATCA 121868
QY      1093 TGCTCTGTGATAACCTGCGTATACAGGCAATGTGCGCGGCAAGGCCCTTCTCGGCTATG 1152
Db      121867 TGAGTGGGATCAACATCCAAAGCAACGCGATCTGGCTAAGCGTTTCTCTCGCCTTCG 121808
QY      1153 CCAAGGCGCGGATCGGAGTTGGGAGTGGGATTTAGGAAACGCGACCTTCCGGAAG 1212
Db      121807 CACATTCGCTGCTTCTGAGCTCGCGAATGGGTGGAAACACCTGGCTTCCCAACT 121748
QY      1213 GAATGGTTGATCGCATACCCCGACCGTTTCGGCGGAAATGCCAAGAGCTCAACGGGG 1272
Db      121747 CCATGGTGGACCGCATCACCC---CTGAAACACCGACCGCGCATGACATCAAGG 121691
QY      1273 CCAGTGGGCTGATCAGCAACCTGCGCGTGGTGGCGGAGATTTCCATCAGTGGGTGCTGG 1332
Db      121690 AAATCGGCTACATCATCGTGGCGGAGTGGTTCTGAAGATTTCAACCAATGGGTCCTCG 121631
QY      1333 AAGACAGTTTCCGATGCGGCTGCGCGCTTGAAGAAACGCGGTGCGAGATGGTGGGG 1392
Db      121630 AGGATGCTTTCACCCAGGCGCCCGCGTACGAGAGGTGGCGGTGCGAGTCTCTCG 121571
QY      1393 ACGTGACGAGTGGGAGTACGTCAGATCCGAATCTCAATGACGGGATGTCATGCTCT 1452
Db      121570 ACGTGGAGCTTATGAATTAATGAAGTGGCGCTGCTCAAGCGCTCCACACGAGCATTT 121511
QY      1453 GCTTCCAGGCAATCTGCTGGCTATGAGATGGATGAGCCATTCAGACACGCAAC 1512
Db      121510 GCTACTTCCGCACTTGGCTGGCGACCATGTCACAGAGTGTCCAGCGATCATGCCGATCCGCT 121451
QY      1513 TCTTTGGCAATCTGAAGAATATCTCAACAGAGTGTATCCGACCGCTGAGGCGCTT 1572
Db      121450 FCCAGATTTCTCTGCTTACATGGAGCGGGAAGCACCCCTACCCTCAAGGAACATTC 121391
QY      1573 CAGGATGACGCTCGAAGGCTATCGGACAGCGCTCATCAGCGGTTTCTCCAAACAGGGA 1632
Db      121390 CAGGTGTGATCTAGATGCTTATGACGCCAACTCATCGCGGATTCGCAACGCGCAG 121331
QY      1633 TGTCGACACGAGCTCGGATGCTAGGATGCTTCCAGGTTCAGGTGCTTCTGGA 1692
Db      121330 TCAAGACACCGTACCGCGCTGTGTGGGAATCTCCGACCGCATTCCAAAGTGGCTGT 121271
QY      1693 CGGAACCGTGTGCGGCGATCGAAGACAGCGGACCTGTACGATATAGGTTCGGAA 1752
Db      121270 TGCCAGTGTACGCGAAGAACTTCGACGAGCGCGGCGAGTCACTTTCTGACGCCATCG 121211
QY      1753 TTGATCTCTATCTCGAAATGCTGCGTGGTTCGCGAGGAGGCG 1796
Db      121210 TCGCATCTGGGCGGCTACGAGAGGACCGGACCGGCGGCG 121167

```

RESULT 3

```

US-09-802-208B-1
; Sequence 1, Application US/09802208B
; Publication No. US20030041352A1
; GENERAL INFORMATION:
; APPLICANT: Parriott, Wayne
; APPLICANT: LaFayette, Peter
; APPLICANT: Kane, Patrick
; TITLE OF INVENTION: Arabitol or Ribitol As Positive Selectable Markers
; FILE REFERENCE: USA-855R
; CURRENT APPLICATION NUMBER: US/09/802,208B
; CURRENT FILING DATE: 2001-03-08
; NUMBER OF SEQ ID NOS: 5
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 1
; LENGTH: 1848
; TYPE: DNA

```

```

; ORGANISM: Escherichia coli
US-09-802-208B-1
Query Match      2.6%; Score 105; DB 9; Length 1848;
Best Local Similarity 46.8%; Pred No. 4.5e-22;
Matches 509; Conservative 0; Mismatches 560; Indels 18; Gaps 5;
QY      622 TGCAATTCGGTGTAGTAACTTTTCGAGCCCATAGGCGTCTACGTCGAGCAGATTC 681
Db      503 TGCACATCGGTTAGTGTCTTTTCATCGGCACATCAGCGTGGTATCTACACCGTTTGC 562
QY      682 TTGAACACGCTCCGCACTGGCGGATTTGGTGTGGCTGACGGCGAGTGCACCTTCAA 741
Db      563 AGGTGATGGGCGATAAAGCACTGGAGCATTTGTCGGGCAATATTCTGTAATGA---TGCTG 619
QY      742 AGAAAAAGCCGAGCAATTTCAAGGCCCGAGACTGCTGTATTCTCCCTGACCGAGAGCGGTC 801
Db      620 AACATGTCGTACAGCACTCAGTGCACAGAAAGTGCCTATGCTGTAAGAAACCGTCAGCC 679
QY      802 CGTCGGCAAGACGACGCTGCTGATGGCGGCTGATGGCGGCTGCTGCTGCTGCTGCTGCTG 861
Db      680 CGGAAGGGTAAAGCAATATGAAGAGATCACCTCAATTCAGAAATGATACCGTGGCAGG 739
QY      862 CCGATCCGGAAGCCGCTGCTGAGCATCTTGTGTGATCCGCGCATCCGCTGCTGCTTCAATGA 921
Db      740 CAGATTTACAACCGCTGATTTGCTGAAGGGGCGAGATCCGAAGACAAAGTGTGCTTTCA 799
QY      922 CGATCAGCAAGCGGCTTACAACTCAACGAGACGACCGGTGCTGCTGATCTGGAGATG 981
Db      800 CCGTCAACGAGCGGCTACTACCTGA---ATACCACTCAAACTGGAAGTTTACAAATC 856
QY      982 CGGCACTAAAGCGCGACCTCAAGAACCCGCGGAAAGCGCTCTACCTTTTTCGTTACGTGG 1041
Db      857 CTGATTTAGCGCGAGATCTTAAAGGGGATGCAAAACAATTTACGGTGTATTACCGCTA 916
QY      1042 TCGAGGCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1101
Db      917 TCTTGAAGCGGCTATGCAAAATAAACGCGG-----ACCATAACCTGCTGCTGCTGCTGCTG 970
QY      1102 ATAACCTGCTCATAAACGCAATGTCGCGCGCAAGCGCTTCTGCTGCTGCTGCTGCTGCTG 1158
Db      971 ATAAGTGGCGCAATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1030
QY      1159 CGCGGATCCGAGTGGCGAAGTGGATTTGAGGAAACGCGACCTTCCGCAACGGAATGG 1218
Db      1031 CTGCAAAACAGGATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1090
QY      1219 TTGATCGCATCACCCGACCGTTTCGCGGGAATCGCAAGAGCTCAACGCGGCGCAGT 1278
Db      1091 TTGACCGGATTAACGCTCTGCTGCGGAGAGAACTTCCGCGCAGAGAACTTCCGCGCTCAAACGG 1150
QY      1279 GGCTGGATCAGCACTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1338
Db      1151 GTATTGCGGATTAAGCGCGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1210
QY      1339 AGTTTGGGATGGCGCTCCGCGCTTGAAGCGGCTGAGATGGTGGGAGGAGTGGTGGGAGGAGT 1398
Db      1211 ATTTCCGCTGATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1270
QY      1399 CGGACTGGGAGTACGTCAGATTCGCAATGCTCAATGACGAGGATGCTGCTGCTGCTGCTGCT 1458
Db      1271 TCCCTATGAAGAGGCGAAGATTCGCAATTTCACTTCACTTCACTTCACTTCACTTCACTT 1330
QY      1459 CAGCATTTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1518
Db      1331 CAGGTACCTTTATCGTCAAAAATATATCCAGAAAGCAATGACCAATGACCAATGACCAATG 1390
QY      1519 GCAATCTGAAGAACTATCTCAACAGGATGCTATCCCGACCGCTGAAGCGGCTTCAAGCA 1578
Db      1391 AGATTGCGCGCGCTACGTCAGCGAAGATGCTATTCCTTGTCT---GGCGGATTAACGTA 1447
QY      1579 TGACGCTCGAAGGCTATCGGACAGCGTCACTACGCGCTTCTCCAAAGGCGATGCTCG 1638

```

Db 1448 TCGATTTGCCAACCTACCGTATGTTGTAAGCTTTTACCAATCCACATATTCAGG 1507
 Qy 1639 ACCAGAGCTCCGATGCTAGCATGCTTCCAAAGTTTCAAGTTCAGTCTTCTGGACGGAAA 1698
 Db 1508 ACACCAACCAACGCGCTGCGGATGTTTCTCGAAAAATTCGGCGATGATTCGCCCCA 1567
 Qy 1699 CCGTGGC 1705
 Db 1568 CACTGGC 1574

RESULT 4

US-09-789-561-27
 ; Sequence 27, Application US/09789561
 ; Patent No. US20020064818A1
 ; GENERAL INFORMATION:
 ; APPLICANT: NI et al.
 ; TITLE OF INVENTION: 52 Human secreted proteins
 ; FILE REFERENCE: P2043P1
 ; CURRENT APPLICATION NUMBER: US/09/789,561
 ; CURRENT FILING DATE: 2001-02-22
 ; PRIOR APPLICATION NUMBER: PCT/US00/24008
 ; PRIOR FILING DATE: 2000-08-31
 ; PRIOR APPLICATION NUMBER: 60/152,317
 ; PRIOR FILING DATE: 1999-09-03
 ; PRIOR APPLICATION NUMBER: 60/152,315
 ; PRIOR FILING DATE: 1999-09-03
 ; NUMBER OF SEQ ID NOS: 194
 ; SOFTWARE: PatentIn Ver. 2.0
 ; SEQ ID NO 27
 ; LENGTH: 2371
 ; TYPE: DNA
 ; ORGANISM: Homo sapiens
 ; US-09-789-561-27

Query Match 1.0%; Score 42; DB 10; Length 2371;

Best Local Similarity 50.0%; Pred. No. 0.027;

Matches 105; Conservative 0; Mismatches 105; Indels 0; Gaps 0;

Qy 1157 GCGCGCGATCCGGAGTGGCGAAGTGGATTGAGGAAACCGGACCTTCCCGAACGGAAT 1216
 Db 527 GCGAGCCGGCTGGGCTTTGTCATGCTTCCAGAACCTGGTCTATGATGAGGACTT 586
 Qy 1217 GTTGTATCGCATCACCCGACCGCTTTCGGCGGAAATCCCAAGAGCTCAACGCGGCGAG 1276
 Db 587 TGTGGACTGGTTCATCCCGGACATCCCAAGGACATCAGCCAGCAGATCCACAAAGAGAA 646
 Qy 1277 TGGGTGATGATGACGACCTGCGCTGGTGGCGGAGGATTTCCATCAGTGGTCTGTTGAAGA 1336
 Db 647 GGTGCTCATGTTGGAGCTGTTTCATCGCGGAGGAGCAAGACAGCAGCTGCTGGAAC 706
 Qy 1337 CCAGTTTCGGATGCGCTGCGCGCTTGA 1366
 Db 707 CTGGATGAGAGGAGGAGCGGACGAGGACGA 736

RESULT 5

US-09-822-846-491/C
 ; Sequence 491, Application US/09822846
 ; Publication No. US20030027139A1
 ; GENERAL INFORMATION:
 ; APPLICANT: Jacobs, Kenneth
 ; APPLICANT: McCoy, John M.
 ; APPLICANT: Lavallie, Edward R.
 ; APPLICANT: Collins-Racie, Lisa A.
 ; APPLICANT: Evans, Cheryl
 ; APPLICANT: Merberg, David
 ; APPLICANT: Treacy, Maurice
 ; APPLICANT: Agostino, Michael J.
 ; APPLICANT: Steininger II, Robert J.
 ; APPLICANT: Bowman, Michael R.
 ; APPLICANT: Spaulding, Vikki
 ; APPLICANT: Wong, Gordon G.

; APPLICANT: Clark, Hilary
 ; APPLICANT: Fechtel, Kim
 ; APPLICANT: Howes, Steven H.
 ; APPLICANT: Resnick, Richard J.
 ; APPLICANT: Gulukota, Kamalakara
 ; APPLICANT: Graham, James R.
 ; APPLICANT: Genetics Institute, Inc.
 ; TITLE OF INVENTION: POLYNUCLEOTIDES ENCODING NOVEL SECRETED PROTEINS
 ; FILE REFERENCE: GIN 6400
 ; CURRENT APPLICATION NUMBER: US/09/822,846
 ; CURRENT FILING DATE: 2001-03-29
 ; PRIOR APPLICATION NUMBER: 60/195,605
 ; PRIOR FILING DATE: 2000-04-06
 ; NUMBER OF SEQ ID NOS: 629
 ; SOFTWARE: PatentIn Ver. 2.0
 ; SEQ ID NO 491
 ; LENGTH: 2614
 ; TYPE: DNA
 ; ORGANISM: Homo sapiens
 ; US-09-822-846-491

Query Match 1.0%; Score 41.6; DB 9; Length 2614;

Best Local Similarity 51.3%; Pred. No. 0.037;

Matches 121; Conservative 0; Mismatches 114; Indels 1; Gaps 1;

Qy 3337 CTGATGATCTTTTGTATCAAGCGTCTCCCAATGTCGCAAAAGTGTGAAGTGTG 3396
 Db 2172 CTGCTGCTGCTGTTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 2113
 Qy 3397 CGGGAACGGATCAATAAGTTCCGCGGCGGCTGCTGCTGCTGCTGCTGCTGCTGCTG 3456
 Db 2112 CCGCAGGATCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 2053
 Qy 3457 GTTCTGTCCATCCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 3516
 Db 2052 TTGCTGTTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1993
 Qy 3517 GCGCAGGATATTCGTTTCCGCTGCGCTTCCAGGCGCGCTTTCGCGCTGATCGA 3572
 Db 1992 AGGTAGGATGG-CTGTTGAGCGAGCGCTGCTGCGCGCTTGGGCACTATTGGA 1938

RESULT 6

US-09-822-849A-393
 ; Sequence 393, Application US/09822849A
 ; Patent No. US20020045170A1
 ; GENERAL INFORMATION:
 ; APPLICANT: Wong, Gordon G.
 ; APPLICANT: Clark, Hilary
 ; APPLICANT: Fechtel, Kim
 ; APPLICANT: Agostino, Michael J.
 ; APPLICANT: Howes, Steven H.
 ; APPLICANT: Resnick, Richard J.
 ; APPLICANT: Gulukota, Kamalakara
 ; APPLICANT: Graham, James R.
 ; APPLICANT: Genetics Institute, Inc.
 ; TITLE OF INVENTION: POLYNUCLEOTIDES ENCODING NOVEL SECRETED PROTEINS
 ; FILE REFERENCE: GIN 6403
 ; CURRENT APPLICATION NUMBER: US/09/822,849A
 ; CURRENT FILING DATE: 2001-03-04
 ; PRIOR APPLICATION NUMBER: 60/195,582
 ; PRIOR FILING DATE: 2000-04-06
 ; NUMBER OF SEQ ID NOS: 598
 ; SOFTWARE: PatentIn Ver. 2.0
 ; SEQ ID NO 393
 ; LENGTH: 1512
 ; TYPE: DNA
 ; ORGANISM: Homo sapiens
 ; US-09-822-849A-393

Query Match 1.0%; Score 41.4; DB 10; Length 1512;

Best Local Similarity 48.5%; Pred. No. 0.034;

Matches 114; Conservative 0; Mismatches 121; Indels 0; Gaps 0;


```
Db 560 ACGGTGGCGCTTTAAACAAGGGTTTCATCGATTGGCTTGAAGAGCAAAAGCTGTTCTGCA 619
Qy 1210 ACGGAATGGTTGATCGATCAACCCGACCGCTTTCGGCGGAAATCGCCAAAGAGCTCAACG 1269
Db 620 ATACGCTTGTGGACAGATCTACCGGATTCGGGAATCCCGAAGAANAATGCACACCGCTTGAAG 679
Qy 1270 CGGCAGTGGCTGGATGACGACCTGCGCTGGTGGCGGAGGATTTCCATCAGTGGGTGC 1329
Db 680 AGCGCGCGGCTATCAGCATTCCTTGATGGTGAAGCGGACCGCTATCATTTGTTGCTGA 739
Qy 1330 TGAAGACCACTTTGGCGATGG-----CGTCGCCGCTTGAAGAGCCGG-----CGTGC 1380
Db 740 TTGAAGGCGCGAAATGGCTTCGAGACGAGCTTCGCTTCATGAGCGCGGCTGAATGTG 799
Qy 1381 AGATGGTCGGGACGTGACGCACTGGGAGTACGTCAAGATCCGATCCGATGCTCAATCGAGGC 1440
Db 800 AGTTGTCCAGGATGTGACGCGGTACAGGATGAAGAAAGTGAAGATTTAAACGCGCGCC 859
Qy 1441 ATGTCATGCTCTGCTTCCAGGCAATTCGTGCGGCTATGAGAATGTGGATGACGCCATTG 1500
Db 860 ATACCGCATGACCGCGGTCTCTTATTTGGCAGGCTTGAGACGGTTCCGGAAGCGGTG 919
Qy 1501 AAGACGGAATCTCTTGGCAATCTGAAGAATCTATCTCAACAAGAGTGTATCCCGACCC 1560
Db 920 AAGA---CGAGCTTATCGCTCATTTGTGAAAGAGCTGATCCAGCGGAGGTGCTCGAGA 976
Qy 1561 TGAAGGCGCTTACGAGCATGACGTGCAAGCTATCGGACGCTATCAGCGCTTCT 1620
Db 977 CGATCGATATCCCGATAGCAGGTGACAACTAGTGGGAAGAGATTCGACCGCTTTA 1036
Qy 1621 CCAACAAGCGATGTCGACGACGCTCGGATTCGGATTCGCTGATGCTTCCAAAGTTC 1680
Db 1037 AAACCCGCTTTATCCGTCAACAAGCTGATCGATATTTCTTAAATTCCTTTCCAAAATTC 1096
Qy 1681 AGGT 1684
Db 1097 AAGT 1100

RESULT 10
US-10-040-26/c
; Sequence 26, Application US/10152040
; Publication No. US20030077251A1
; GENERAL INFORMATION:
; APPLICANT: ESCRIOU, NICOLAS
; APPLICANT: VAN DER WERF, SYLVIE
; APPLICANT: VIGNUZZI, MARCO
; APPLICANT: GERBAUD, SYLVIE
; TITLE OF INVENTION: REPLICONS DERIVED FROM POSITIVE STRAND RNA VIRUS
; TITLE OF INVENTION: GENOMES USEFUL FOR THE PRODUCTION OF HETEROLOGOUS
; FILE REFERENCE: 03495.0229-00000
; CURRENT APPLICATION NUMBER: US/10/152,040
; PRIOR FILING DATE: 2002-06-27
; PRIOR FILING DATE: 2001-05-23
; NUMBER OF SEQ ID NOS: 28
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 26
; LENGTH: 8017
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: Synthetic
US-10-152-040-26
Query Match 1.0%; Score 40.4; DB 9; Length 8017;
Best Local Similarity 97.6%; Pred. No. 0.14;
Matches 41; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 1 AAGCTTGCATCCCTGCAGGTGCTAGGATCCCGTTTTT 42
|||||
Db 1 AAGCTTGCATCCCTGCAGGTGCTAGGATCCCGTTTTT 4906

RESULT 12
US-10-152-040-28/c
; Sequence 28, Application US/10152040
; Publication No. US20030077251A1
; GENERAL INFORMATION:
; APPLICANT: ESCRIOU, NICOLAS
; APPLICANT: VAN DER WERF, SYLVIE
; APPLICANT: VIGNUZZI, MARCO
; APPLICANT: GERBAUD, SYLVIE
; TITLE OF INVENTION: REPLICONS DERIVED FROM POSITIVE STRAND RNA VIRUS
; TITLE OF INVENTION: GENOMES USEFUL FOR THE PRODUCTION OF HETEROLOGOUS
; FILE REFERENCE: 03495.0229-00000
; CURRENT APPLICATION NUMBER: US/10/152,040
; PRIOR FILING DATE: 2002-06-27
; PRIOR FILING DATE: 2001-05-23
; NUMBER OF SEQ ID NOS: 28
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 28
; LENGTH: 10417
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: Synthetic
US-10-152-040-28
Query Match 1.0%; Score 40.4; DB 9; Length 10417;
Best Local Similarity 97.6%; Pred. No. 0.16;
Matches 41; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 1 AAGCTTGCATCCCTGCAGGTGCTAGGATCCCGTTTTT 42
|||||
```

```
Db 4872 AAGCTTGCATCCCTGCAGGTGCTAGGATCCCGTTTTT 4831

RESULT 11
US-10-152-040-27/c
; Sequence 27, Application US/10152040
; Publication No. US20030077251A1
; GENERAL INFORMATION:
; APPLICANT: ESCRIOU, NICOLAS
; APPLICANT: VAN DER WERF, SYLVIE
; APPLICANT: VIGNUZZI, MARCO
; APPLICANT: GERBAUD, SYLVIE
; TITLE OF INVENTION: REPLICONS DERIVED FROM POSITIVE STRAND RNA VIRUS
; TITLE OF INVENTION: GENOMES USEFUL FOR THE PRODUCTION OF HETEROLOGOUS
; FILE REFERENCE: 03495.0229-00000
; CURRENT APPLICATION NUMBER: US/10/152,040
; PRIOR FILING DATE: 2002-06-27
; PRIOR FILING DATE: 2001-05-23
; NUMBER OF SEQ ID NOS: 28
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 27
; LENGTH: 8092
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: Synthetic
US-10-152-040-27
Query Match 1.0%; Score 40.4; DB 9; Length 8092;
Best Local Similarity 97.6%; Pred. No. 0.14;
Matches 41; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 1 AAGCTTGCATCCCTGCAGGTGCTAGGATCCCGTTTTT 42
|||||
Db 4947 AAGCTTGCATCCCTGCAGGTGCTAGGATCCCGTTTTT 4906

RESULT 12
US-10-152-040-28/c
; Sequence 28, Application US/10152040
; Publication No. US20030077251A1
; GENERAL INFORMATION:
; APPLICANT: ESCRIOU, NICOLAS
; APPLICANT: VAN DER WERF, SYLVIE
; APPLICANT: VIGNUZZI, MARCO
; APPLICANT: GERBAUD, SYLVIE
; TITLE OF INVENTION: REPLICONS DERIVED FROM POSITIVE STRAND RNA VIRUS
; TITLE OF INVENTION: GENOMES USEFUL FOR THE PRODUCTION OF HETEROLOGOUS
; FILE REFERENCE: 03495.0229-00000
; CURRENT APPLICATION NUMBER: US/10/152,040
; PRIOR FILING DATE: 2002-06-27
; PRIOR FILING DATE: 2001-05-23
; NUMBER OF SEQ ID NOS: 28
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 28
; LENGTH: 10417
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: Synthetic
US-10-152-040-28
Query Match 1.0%; Score 40.4; DB 9; Length 10417;
Best Local Similarity 97.6%; Pred. No. 0.16;
Matches 41; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 1 AAGCTTGCATCCCTGCAGGTGCTAGGATCCCGTTTTT 42
|||||
```

[illegible]

OTHER INFORMATION: ORF 24; negative strandedness
NAME/KEY: misc.feature
LOCATION: (79864)..(78107)
OTHER INFORMATION: ORF 25; negative strandedness
NAME/KEY: misc.feature
LOCATION: (81624)..(79861)
OTHER INFORMATION: ORF 26; negative strandedness
NAME/KEY: misc.feature
LOCATION: (81909)..(81682)
OTHER INFORMATION: ORF 27; negative strandedness
NAME/KEY: misc.feature
LOCATION: (82346)..(82062)
OTHER INFORMATION: ORF 28; negative strandedness
NAME/KEY: misc.feature
LOCATION: (82587)..(84446)
OTHER INFORMATION: ORF 29; positive strandedness
NAME/KEY: misc.feature
LOCATION: (84481)..(85548)
OTHER INFORMATION: ORF 30; positive strandedness
NAME/KEY: misc.feature
LOCATION: (85556)..(86845)
OTHER INFORMATION: ORF 31; positive strandedness
NAME/KEY: misc.feature
LOCATION: (87372)..(86803)
OTHER INFORMATION: ORF 32; positive strandedness
NAME/KEY: misc.feature
LOCATION: (87494)..(88420)
OTHER INFORMATION: ORF 33; positive strandedness; N-terminus only
US-09-976-059-1

Query Match 1.0%; Score 40.2; DB 9; Length 88421;
Best Local Similarity 50.8%; Pred. No. 0.45;
Matches 96; Conservative 0; Mismatches 93; Indels 0; Gaps 0;
QY 3445 ATGCTGGCGGGTCTCTCCATCTCGCTGTCTGTGGGTCTGTGGAGCTGCCGGTCGGC 3504
Db 37184 ACGGTCCGGCGAGCGGATCTCCGGCGCATCCGGCGAGGTGTCCGCCAGTAGGC 37125
QY 3505 GTGATGTTGCTGCGCAGATATTCGGTCTTCGGTCGCCTTCAGGCGCGCTCTTGGC 3564
Db 37124 GAGCTGGCGCTCGCGCAGGCTCTCCGGGTACCGCGCGCGCGCCAGCAACTCTCTCTGCCA 37065
QY 3565 TGGATCGAAGCTCAACATCCGATGCTGGGCGCTTCGGCGGAAAGCGGAGAGCTAA 3624
Db 37064 GAGCGCGTAGTCGGGGTACTCAGCGCGCAGCGGCTCCCGAGCGGTGCCCGCGGCCAG 37005
QY 3625 CCGTTCGTC 3633
Db 37004 CCGGCGCTC 36996

RESULT 15
US-09-981-353-85/c
Sequence 85, Application US/09981353
Patent No. US20020160382A1
GENERAL INFORMATION:
APPLICANT: Lasek, Amy W.
APPLICANT: Jones, David A.
TITLE OF INVENTION: GENES EXPRESSED IN COLON CANCER
FILE REFERENCE: PA-0038 US
CURRENT APPLICATION NUMBER: US/09/981,353
CURRENT FILING DATE: 2001-10-11
NUMBER OF SEQ ID NOS: 194
SOFTWARE: PERL Program
SEQ ID NO 85
LENGTH: 3115
TYPE: DNA
ORGANISM: Homo sapiens
FEATURE:
NAME/KEY: misc.feature
OTHER INFORMATION: Incyte ID No. US20020160382A1 1102315.3
NAME/KEY: unsure
LOCATION: 2713, 2719, 3094

OTHER INFORMATION: a, t, c, g, or other
US-09-981-353-85
Query Match 1.0%; -Score 40; DB 9; Length 3115;
Best Local Similarity 46.4%; Pred. No. 0.13;
Matches 130; Conservative 0; Mismatches 150; Indels 0; Gaps 0;
QY 2628 GAACAGTCCCGTCATGAGGTTCTCAAGCGGCGCGTATTATCGGCATAGGCTTGGCCAT 2687
Db 773 GCACAGGTGGTAGATGAAGGACCCAGGCGCTCTTGTCACTGACATGCTCCGTGACCCAC 714
QY 2688 TTCGGGGCATACATCTCGAAATCGTCCGTCGGCGCGGATCGAACAGCATGCC 2747
Db 713 CATCTCCTCTGGATGATGTAGCTGCGGCAGGTAGTCCCGCTCTTTCACGTTTCATGAG 654
QY 2748 GACTTCTTGTGTATTCGGGGGGAACCTGGAAGCAGGTCTTGAAGCGTTGATTTCTGTG 2807
Db 653 GAGCTCCAGAAAGTTGCGAGGGGCGAGCACAATGTTGTTGATTTGATTCGATGACATAGCA 594
QY 2808 TCGGTTCAACGCGCGCTTCGCCAGCTTCGCCGACAGGTCAGACCCCGCTGGAAGTCATGATGATGTC 2867
Db 593 CTTGTCCAGGAGATATCATGTAGCAGTCCAGACCCCGCTGGAAGTCATGATGATGTC 534
QY 2868 GTAAAGCTGATCTCGTTTTCGCCAGGCGCGCAGCAATCTTG 2907
Db 533 TGCAGGCTCACCGCGCGCAAACTGGGGCACAGGCACGTTG 494

Search completed: May 18, 2003, 21:32:28
Job time : 3984 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: May 18, 2003, 21:32:35 ; Search time 5382 seconds
(without alignments)
12382.837 Million cell updates/sec

Title: US-09-926-163B-1
Perfect score: 4115
Sequence: 1 aagtttcacgtcctgcaggt.....tcggggcgccctgaagctt 4115

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 16154066 seqs, 8097743376 residues

Total number of hits satisfying chosen parameters: 32308132

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

EST:*

1: em_estba:**

2: em_esthum:**

3: em_estin:**

4: em_estmu:**

5: em_estov:**

6: em_estpl:**

7: em_estro:**

8: em_htc:**

9: gb_est1:**

10: gb_est2:**

11: gb_htc:**

12: gb_est3:**

13: gb_est4:**

14: gb_est5:**

15: em_estfun:**

16: em_estom:**

17: gb_gss:**

18: em_gss_hum:**

19: em_gss_inv:**

20: em_gss_pln:**

21: em_gss_vrt:**

22: em_gss_fun:**

23: em_gss_mam:**

24: em_gss_mus:**

25: em_gss_other:**

26: em_gss_pro:**

27: em_gss_rod:**

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
C 1	180.8	4.4	801	17	BH395246
C 2	155.4	3.8	761	17	AZ139030
C 3	146.8	3.6	840	10	BE641454
C 4	131.8	3.2	799	17	CNS010W5
5	88.6	2.2	527	9	AA965891
6	86.6	2.1	481	17	AZ049105

7	85.4	2.1	624	10	AW671891	AW671891 LG1_352-G
8	80.4	2.0	519	13	BI191648	BI191648 11e03fs.r
9	80.4	2.0	519	13	BI191768	BI191768 12e03fs.r
10	77.6	1.9	520	17	BH143537	BH143537 TDGEB13TH
11	69.4	1.7	556	17	AQ500091	AQ500091 V34C6 mTn
12	54.2	1.3	506	12	BG608411	BG608411 296895 MA
13	51.4	1.2	788	17	AZ933800	AZ933800 BJ_Ba000
14	49.2	1.2	1101	17	CNS017SY	AL108460 Drosophil
15	48	1.2	861	9	AL539916	AL539916 AL539916
16	47.6	1.2	927	12	BE733079	BE733079 601569589
17	47.2	1.1	244	17	AZ578838	AZ578838 28a07 Sho
18	47	1.1	830	14	BQ672858	BQ672858 AGENCOURT
19	47	1.1	934	14	BQ641340	BQ641340 AGENCOURT
20	47	1.1	1012	13	BM549784	BM549784 AGENCOURT
21	46.4	1.1	812	13	BI600315	BI600315 603250632
22	46.4	1.1	831	12	BG818263	BG818263 602779895
23	46.4	1.1	861	14	BQ437338	BQ437338 AGENCOURT
24	46.4	1.1	875	13	BM009512	BM009512 603630050
25	46.4	1.1	882	13	BI856423	BI856423 603387791
26	46.4	1.1	898	12	BG470953	BG470953 602511926
27	46.4	1.1	922	12	BG754237	BG754237 602709788
28	46.4	1.1	925	17	CNS0091P	AL053013 Drosophil
29	46.4	1.1	992	14	BQ057936	BQ057936 AGENCOURT
30	46.4	1.1	1016	13	BM562009	BM562009 AGENCOURT
31	46.2	1.1	824	12	BG831754	BG831754 602766057
32	46.2	1.1	1030	13	BM563059	BM563059 AGENCOURT
33	46	1.1	121	17	AZ579022	AZ579022 30b10 Sho
34	46	1.1	912	14	BQ950903	BQ950903 AGENCOURT
35	45.8	1.1	897	14	BQ933488	BQ933488 AGENCOURT
36	45.6	1.1	630	13	BQ272413	BQ272413 BJ272413
37	45.6	1.1	925	17	CNS0091P	AA383457 EST96834
38	45.4	1.1	221	9	AA383457	AA383457 EST96834
39	45.4	1.1	770	12	BG756596	BG756596 602713747
40	45.4	1.1	788	12	BG756729	BG756729 602715595
41	45.4	1.1	885	13	BM048024	BM048024 603620340
42	45.2	1.1	681	10	BB643958	BB643958 BB643958
43	45.2	1.1	786	9	AI929012	AI929012 au64all.y
44	45.2	1.1	864	14	BQ231485	BQ231485 AGENCOURT
45	44.8	1.1	765	12	BG764520	BG764520 602736565

ALIGNMENTS

RESULT 1
BH395246/C
LOCUS BH395246 801 bp DNA linear GSS 11-DEC-2001
DEFINITION AG-ND-144F15.TR ND-TAM Anopheles gambiae genomic clone AG-ND-144F15
ACCESSION BH395246
VERSION BH395246.1 GI:17341387
KEYWORDS GSS.
SOURCE African malaria mosquito.
ORGANISM Anopheles gambiae
Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
Neoptera; Endopterygota; Diptera; Nematocera; Culicoidea;
Anopheles.
REFERENCE 1 (bases 1 to 801)
AUTHORS Shetty,J., Malek,J., Koo,H., Collins,F., Gardner,M. and Loftus,B.J.
TITLE Direct Submission of BAC-end sequences from Anopheles gambiae
JOURNAL Unpublished (2001)
COMMENT Other_GSS: AG-ND-144F15.TF
Contact: Brendan J Loftus
Department of Eukaryotic Genomics
The Institute for Genomic Research
9712 Medical Center Dr., Rockville, MD 20850, USA
Tel: 301 838 0208
Fax: 301 838 3543
Email: bjloftus@tigr.org
This clone is from an A. gambiae BAC library (ND-TAM) provided by
F.H. Collins and sequenced by The Institute for Genomic Research
(TIGR). The BAC library was generated from A. gambiae PEST strain
DNA. All DNA was extracted from newly hatched first instar larvae


```
Db 315 GTTCGGACGCTCGGAATGTAAGGAAACCGGAGCCGAACCTGGTAGCGATGTCCTGCCT 256
QY 1404 TGGAGTAGCTCAAGATCCGAATGCTCAATACGAGGATGTCATGCTCTGCTTCCAGGC 1463
Db 255 TATGAAGAGATGAAGTTGGCGATGCTCAACGGCAGTATTCATTCCTGGCGTATCTGGG 196
QY 1464 ATTCTGGTGGCTATGAGATGTGGATGACGCCATTTGAAGA 1504
Db 195 TATCTTCAGGATATACGACATTAATGACTGTATGAAGA 155

RESULT 3
BE641454 840 bp mRNA linear EST 01-SEP-2000
LOCUS Ceratopteris Spore Library Ceratopteris richardii
DEFINITION cDNA clone Cr12_3_F06 5', mRNA sequence.
ACCESSION BE641454
VERSION BE641454.1 GI:9959119
KEYWORDS EST.
SOURCE Ceratopteris richardii.
ORGANISM Ceratopteris richardii.
REFERENCE Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Filicophyta; Filicopsida; Filicales; Pteridaceae; Ceratopteris.
1 (bases 1 to 840)
Chatterjee,A., San Miguel,P., Stout,S.C., Banks,J. and Roux,S.J.
Expressed sequence tags of cDNA clones from a C. richardii library
Unpublished (2000)
Contact: Roux SJ
Section of Molecular Cell and Developmental Biology
University of Texas
Biology Building, Room 16, Austin, TX 78712, USA
Tel: 512 471 4238
Fax: 512 232 3402
Email: sroux@uts.cc.utexas.edu
Plate: Cr12.3 row: F column: 06
Seq primer: Sp6.
Location/Qualifiers
1. 840
/organism="Ceratopteris richardii"
/cultivar="Brogn"
/db_xref="taxon:49495"
/clone="Cr12_3_F06"
/clone_lib="Ceratopteris Spore Library"
/tissue_type="Gametophyte"
/cell_type="Spore"
/dev_stage="20 hours after germination initiation"
/note="vector: pCMVSPORT6; EST sequence from cDNA library.
cDNA library constructed from mRNA isolated from C.
richardii spores that had developed for 20 hours after
their germination had been initiated by white light."
BASE COUNT 168 a 313 c 226 g 133 t
ORIGIN

Query Match 3.68; Score 146.8; DB 10; Length 840;
Best Local Similarity 50.68; Pred. No. 2e-27;
Matches 406; Conservative 0; Mismatches 392; Indels 4; Gaps 2;

QY 576 GTCCAGGCTCCCTATGACATGACGCGGATCAAGCCTGGGATCGTCATTTGGGTGTA 635
Db 43 GTCAAAATCCGACCTACGGCAGGATGACGTACGAGGGAATGTCCAGTCGGGCTC 102
QY 636 GGTAACCTTTTTCGAGCCCATGAGCGCTTCTACGTGAGCAGATTTTGAACAGCTCCG 695
Db 103 GGCGGCTTCCACAGAGCCCATCTGGCGCTCTACGTCACAAATTTGATGCAACACCGC 162
QY 696 GACTGGCGGATTTGTTGGTGTGGCTACGGGCGAGTACCGTTCAAGAAAGGCGCAG 755
Db 163 CACCGCGACTGGGCGCATCTGGCGGCTGGCGTGGACCCCAACGACGCGGCATCGCGAC 222
QY 756 GAATTAAGGCCCGAGGACTGCTTATTCCTGACCGAGCGGCTCCGTCGGCAAGAGC 815
Db 223 GTGCTCAGGCCCGAGGACCACTGTACACCGTATCAGCGGCTCCGCAAGGGC---AGC 279
```

```
QY 816 ACGTGCCTGCTATGGCGCGCTGCGTACTATCTGCTTGCCTCCCGCGCGATCCGGAAGCC 875
Db 280 TTCGCCAACGCTGTCGCGAGCATCAACTGCTTCTTTCGCCCGCGAGCAGCGGAGGCC 339
QY 876 GTGCTGAAGCATCTTGTGATCGCGCATCGGATCGTTTCCATGAGCATCAGGAGGC 935
Db 340 GTCATTGCCAAGATGGCGCACCCCGACACCATCTCTCCCTGAGCATCAGGAGAGC 399
QY 936 GGCTACAACTCAACGAGAGCAGCGGTTCGATCTGGAGAATGGCGGAGTAAAGGCC 995
Db 400 GGCTACTACTACACGAAACACCCACAGCTGAGGCCAACCATTCGGACATCCAGCAC 459
QY 996 GACCTCAAGAACCCGAAAAAGCCGCTTACCGTTTTCGTTTACGTTGGTGGAGCCCTCGGT 1055
Db 460 GACCTGGCCAAAGAGAACGCCCGCTCAGCACCTTTGGCTTCTTACGCCGCGCTGGCC 519
QY 1056 CGTCTTGGGATGCGCGTGTAAAGCATTTACGGTCATGCTCTGTGATACCTCGGTCAT 1115
Db 520 AAGCGCCACGCGAAGGCCTTCAAGCCCTTCAAGCCCTTCAAGCCCTTCAAGCCCTTCAAG 579
QY 1116 AACGGCAATGTGCGCCCGCAAGGCCTTCTCGGCTATGCGAAGCGCGCGATCCGGAGTTG 1175
Db 580 AAGCGGTCCATCACCCCGCACATGCTCGAGTCGTTTGGCCCGGTCCCAACCCGAGCTG 639
QY 1176 GCGAAGTGGATTGAGGAAACCGGACCTTCCGAAACGGAATGTTGATCGCATCACCCTG 1235
Db 640 GCC-AGTGGATGCGCCAGGAGGTGCTTCCCAACACCTTGGGCGCAGCCGATACACCC 698
QY 1236 ACCGTTTGGGGAAATCGCCAAAGAGCTCAACCGCGCAGTGGGCTGGATGAGCAGCTG 1295
Db 699 CGCACCTTCCAAAACGACCTTTGAGTCTTGGCCAGAGCTTTGGCCCTCGAGGAGCGCTGG 758
QY 1296 CCGCTGTGCGCGAGGATTTTCATCATGCTGGTGTGGAAGACCACTTTGCGGATGGCGCT 1355
Db 759 CCGCTTGTACCCGAACCTTCATGTCATGCGGTCTTTGAGGAGACASTTCAGGAGGCGCC 818
QY 1356 CCGCGCTTGAAGAACCGCGG 1377
Db 819 CCGCGCTTCAAGAGGTGGGG 840

RESULT 4
CNS01QW5 799 bp DNA linear GSS 14-JUN-2001
LOCUS Anopheles gambiae GSS SP6 end of clone 31M20 of NotreDamel library
DEFINITION from strain PEST of Anopheles gambiae (African malaria mosquito),
genomic survey sequence.
ACCESSION AL156645
VERSION AL156645.1 GI:7017564
KEYWORDS GSS.
SOURCE African malaria mosquito.
ORGANISM Anopheles gambiae
Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
Neoptera; Endopterygota; Diptera; Nematocera; Culicoides;
REFERENCE 1 (bases 1 to 799)
AUTHORS Direct Submission
TITLE Genoscope.
JOURNAL Submitted (16-FEB-2000) Genoscope - Centre National de Sequencage :
BP 191 91006 EVRY cedex - FRANCE (E-mail : seqref@genoscope.cns.fr
- Web : www.genoscope.cns.fr)
2 (bases 1 to 799)
AUTHORS Roth,C.W., Brey,P.T., Ke,Z., Collins,F.H. and Weissenbach,J.
TITLE Direct Submission
JOURNAL Submitted (16-FEB-2000) BBMI, Institut Pasteur, 25, rue du Dr.
Roux, Paris 75015, France
COMMENT This clone is from an A. gambiae BAC library provided by F.H.
Collins and sequenced by Genoscope in collaboration with the
Laboratory of Biochem. and Biol. Molec. of Insects, Institut
Pasteur. Location/Qualifiers
1. 799
source
```

/organism="Anopheles gambiae"
/strain="PEST"
/db_xref="taxon:7165"
/clone="31M20"
/clone_lib="Notredame1"
/note="end : SP6"

BASE COUNT 151 a 245 c 206 g 167 t 30 others
ORIGIN

Query Match 3.2%; Score 131.8; DB 17; Length 799;
Best Local Similarity 51.6%; Pred. No. 1.8e-23;
Matches 332; Conservative 6; Mismatches 302; Indels 4; Gaps 3;

```
QY 861 GCCGATCCGAGCGCTGTAAGCATCTTGTATCGCGCATCCGCATCGTTCCATG 920
DB 649 GACGGCTCGAGACCGCTCTGGCCGATGTGGNACCGCAGGTGGCGATGTGCGTG 590

QY 921 ACGATCACGAGGCGGTACAAATCAACAGACGACCGGTGCTTCGATCTGGAGAA 980
DB 589 ACCATCACGAGAAAGGCTACTGCCATTCGCCGCGACCGGCAACTGATCGATCAC 530

QY 981 CGGCAGTAAGCGCGACCTCAAGNACCGGAAAGCGCTTACCGTTTTCGTTACGTG 1040
DB 529 CTTTATCGTGCAGACCTGMAAACCCGACACCGGAAATCTGCCCGGGTGTGTC 470

QY 1041 GTCGAGCGCCTCGCTGCTGGTGGTAAAGCATTTACCGTTCATGTCCTGT 1100
DB 469 GTGAAAGSCTCGCGCGGTAAAGCGCGCTGCC--GGGTTTAGCTGATGTCCTGC 411

QY 1101 GATACTGCGTATACGGAATGTGCCCGCAGGCTTCTCGGTATGCCAAGCG 1160
DB 410 GACAACATGCCGAGAACGGTCACGTGATGCGMAATGTCACTGCGCGTACGCCGCG 351

QY 1161 CGCATCCGAGTGGCGAAGTGTTCAGGAACGCGACCTTCCCGAAGGATGTT 1220
DB 350 GTAGACGTGAGCTGGCGAGTGGATCGAATCCACGTTTACCTTCCCGTCAACGATG 291

QY 1221 GATCCATCACCCGACCGTTCGCGGAAATCGCAAGAGCTCAACCGCGCGAGTGG 1280
DB 290 GATCGCATNGTGT--CGTCAACCCCGACGCTGGATAAATGAAGCTTACCGGC 232

QY 1281 CTGATGACGACCTGCGCTGCGCGAGGATTTCCATGATGGTGTCTGGAAGACAG 1340
DB 231 GTGCGGATCGCGCGCGCTGTCGAAAGCGCGTGCAGATGTCGCGGAGTGCAGC 1400

QY 1341 TTGCGGATGCGCGCGCTGCAAAAGCGCGTGCAGATGTCGCGGAGTGCAGC 1400
DB 171 TTGTCGCGCGCTGCGCGTGGGACAAAGCGCGAGCTGGTGTGACAGTGCATC 112

QY 1401 GACTGGGAGTACGTCAAGATCCGAATGCTCAATGCGGCGATGTCATGCTTCCCA 1460
DB 111 CCGTTTGAAGATGA--ATGCGATGCTTAACGCGACCCACTMATTCTTGGCTACCTG 54

QY 1461 GCATTCGTGCGGTATGAGAATGTGATGACGCCATTGAAGA 1504
DB 53 GGCTATCTGGGGGCTATCAGMATATTAACGAATGTATGAAGA 10
```

RESULT 5
AA965891
LOCUS o8g06al.r1 Aspergillus nidulans 24hr asexual developmental and vegetative cDNA lambda zap library Emericella nidulans cDNA clone o8g06al 5', mRNA sequence.
DEFINITION AA965891
ACCESSION AA965891
VERSION AA965891.1 GI:3139775
KEYWORDS EST.
SOURCE Emericella nidulans.
ORGANISM Emericella nidulans.
REFERENCE 1 (bases 1 to 527).
AUTHORS Kupfer,D., Gray,J., Hausner,J., Lai,H., Martin,W., Aramayo,R.,

TITLE
JOURNAL
COMMENT

Prade,R. and Roe,B.
An Aspergillus nidulans EST Database
Unpublished (1998)
Contact: Bruce A. Roe, University of Oklahoma, broe@ou.edu
Department of Chemistry and Biochemistry
Advanced Center for Genome Technology, University of Oklahoma
620 Parrington Oval, Norman, OK 73019, USA
Tel: 405 325 4912
Fax: 405 325 7762
Email: broe@ou.edu
We anticipate the future release of the cDNA clones to the Fungal Genetics Stock Center
Seq primer: T3
High quality sequence stop: 405.

FEATURES
source

1..527
Location/Qualifiers
/organism="Emericella nidulans"
/strain="FGSC A26"
/db_xref="taxon:162425"
/clone="o8g06al"
/clone_lib="Aspergillus nidulans 24hr asexual developmental and vegetative cDNA lambda zap library"
/tissue_type="vegetative mycelia, asexual structures"
/note="vector: pBluescript SK-; Site.1: EcoRI; Site.2: XhoI; 5' end of cDNA cloned into EcoRI site of pBluescript 3' end of cDNA cloned into XhoI site of pBluescript"

BASE COUNT 112 a 187 c 128 g 100 t
ORIGIN

Query Match 2.2%; Score 88.6; DB 9; Length 527;
Best Local Similarity 52.3%; Pred. No. 3.9e-12;
Matches 220; Conservative 0; Mismatches 198; Indels 3; Gaps 1;

```
QY 583 CTCCTCCCTATGACATCGACGGGATCAAGCCTGGATCGTGCATTCGGTGTAGTACT 642
DB 39 CTACCTACCCCTCGCAATGCGCGCTCAAGAGGTATTGTCCATATCGCGCGCTCG 98

QY 643 TTTTTCGAGCCATGAGGCGTTCTACGTCGAGCAGATTTTGAACACGCTCCGCACTGG 702
DB 99 TCCACCGAGCCATTTGCGCGTCTATATCGACCGGTGATCGAGNACCACGCGTCACTG 158

QY 703 CGATTGTTGTTGGCTGACGGCGATGACCGTTCAAAGAAAACCCGAGGAATCA 762
DB 159 ACTATGCGATCGCGGCTGTGCTGTGACCCCTTTGACAAGAAGATGCGCATATCTCA 218

QY 763 AGGCCCCAGACTCGCTGTTTCCGTGACCGAGACGGTCCGTCGCGCAAGACGCGTGC 822
DB 219 GGTCTCAAGACCACTCTACACCGTCAATGAGGATGCGGCAAGGC---AGCTTCGCA 275

QY 823 CGCTCATGGCGCGTGGTGAATGCTATGCTTGGCCCCGCGGATCCGGAAGCCGCTGTA 882
DB 276 ACGTGGTGGCTCGATCAACTCGTTCTCTTTGCTGACGACCGCGGAGCCGCTGCTG 335

QY 883 AGCATCTGTTGATCCGCGCATTCGCGATCGTTTCCATGACGATCAGGAAGCGGCTACA 942
DB 336 CCAAGATGGCCACCGCGGACACACATCGTCTCCCTGACTATCACCAATCCCGCTACT 395

QY 943 ACATCAAGCAGACGCGGTGGTTCGATCTGGAGATGCGGCAAGTAAAGCCGACCTCA 1002
DB 396 ACTACATGAGAACACGACGACGATCGTGAAGCGGACCCCGGACATCCAGTTTGACCTG 455

QY 1003 A 1003  
DB 456 A 456
```

RESULT 6
AZ049105
LOCUS
DEFINITION GSSBRU0507 Sheared genomic library Brucella melitensis biovar Abortus genomic clone B4G18, DNA sequence.
ACCESSION AZ049105
VERSION AZ049105.1 GI:7273020

AZ049105 481 bp DNA linear GSS 01-SEP-2000
GSSBRU0507 Sheared genomic library Brucella melitensis biovar
Abortus genomic clone B4G18, DNA sequence.

[illegible]

RESULT 8

BI191648 519 bp mRNA linear EST 10-JUL-2001
 LOCUS l1e03fs.r1 Fusarium sporotrichioides Tri 10 overexpressed cDNA
 DEFINITION library Fusarium sporotrichioides cDNA clone l1e03fs 5', mRNA
 sequence.
 ACCESSION BI191648 GI:14665327
 VERSION EST.
 KEYWORDS Fusarium sporotrichioides.
 SOURCE Fusarium sporotrichioides.
 ORGANISM Eukaryota; Fungi; Ascomycota; Pezizomycotina; Sordariomycetes;
 Hypocreates; Mitosporic Hypocreales; Fusarium.
 REFERENCE 1 (bases 1 to 519)
 AUTHORS Ren,Q., Tag,A., Peplow,A., Lai,H., Kupfer,C., Peterson,A., Beremand
 ,M. and Roe,B.
 TITLE Analysis of a Fusarium sporotrichioides EST database
 JOURNAL Unpublished (2001)
 COMMENT Contact: Bruce A. Roe, University of Oklahoma, broe@ou.edu
 Department of Chemistry and Biochemistry
 Advanced Center for Genome Technology, University of Oklahoma
 620 Parrington Oval, Norman, OK 73019, USA
 Tel: 405 325 4912
 Fax: 405 325 7762
 Email: broe@ou.edu
 Contact Dr. Marian Beremand regarding clone availability Included
 is the best homolog from a blastx search of Genbank nr 04-09-01
 455 5e-45 gi1805056|emb|CAA604 (X86790) N3810 [Saccharomyces
 cerevisiae]
 Seq primer: T3
 High quality sequence stop: 423.
 Location/Qualifiers
 1..519
 /organism="Fusarium sporotrichioides"
 /strain="Tri 10"
 /db_xref="taxon:5514"
 /clone="l1e03fs"
 /clone_lib="Fusarium sporotrichioides Tri 10 overexpressed
 cDNA library"
 /note="Vector: pBluescript SK-; Site_1: EcoRI; Site_2:
 XhoI; 5' end of cDNA cloned into EcoRI site of pBluescript
 XhoI; 3' end of cDNA cloned into XhoI site of pBluescript"
 BASE COUNT 137 a 166 c 116 g 100 t
 ORIGIN
 Query Match 2.0%; Score 80.4; DB 13; Length 519;
 Best Local Similarity 50.9%; Pred. No. 5.7e-10;
 Matches 217; Conservative 0; Mismatches 206; Indels 3; Gaps 1;
 QY 576 GTCCAGGTCCTCCCTATGACATCGACGGGATCAAGCCTGGGATCGTGCATTCGGTGA 635
 DB 97 GTCCAGGTCCTCCCTATGACATCGACGGGATCGTGCATTCGGTGA 635
 QY 636 GACTGGCGATGTTGGTGTGGCTGACGCGGATCGCTTCAAGAAAGGCGGAG 755
 DB 217 GAGCGGATGGCGCATCGCGGATCGCGCTCTCTAAGATGCTGCCATGCCGAT 276
 QY 756 GAATTCAGGCGCCAGGACTGCTGTATTCCTGACCGAGAGCGCTCCGCGCAAGAGC 815
 DB 277 GTCTCTCAAGCCCAAGACCTCTACACCGCTAATTTGAGCGATCGCGAAGGCGAGTTTC 336
 QY 816 ACGGTGCGGTCATCGGCGCGCTGCTGACATATCTGTTCGCCGCGGATCCGGAAGCC 875
 DB 337 GCCGA---CGTTGAGGAGATATACCTCTCTTCAATTCCTCCGACGACCGTGAGGCC 393
 QY 876 GTGCTGAAGCATCTTTGTCGCGCCATCCGCGATCGTTCATGACGATCACGGAAGCC 935
 DB 394 GTCAATGCAAGATGGCTCATCCCGACACCTACATCGTCTCTCTCACTATTACTGAGAGC 453
 QY 936 GGCTACACATCAACGAGACGCCGGTGGTTCGATCTGGAGATGCGGCAGTAAAGGCC 995

DB 454 GGCTACTACTACAGAGAACACCCATCAATTGAAGGATGAGCCCGGACATCCAGCAC 513
 QY 996 GACCTC 1001
 DB 514 GATCTC 519
 RESULT 9
 LOCUS BI191768 519 bp mRNA linear EST 10-JUL-2001
 DEFINITION l2e03fs.r1 Fusarium sporotrichioides Tri 10 overexpressed cDNA
 library Fusarium sporotrichioides cDNA clone l2e03fs 5', mRNA
 sequence.
 ACCESSION BI191768 GI:14665447
 VERSION EST.
 KEYWORDS Fusarium sporotrichioides.
 SOURCE Fusarium sporotrichioides.
 ORGANISM Eukaryota; Fungi; Ascomycota; Pezizomycotina; Sordariomycetes;
 Hypocreates; Mitosporic Hypocreales; Fusarium.
 REFERENCE 1 (bases 1 to 519)
 AUTHORS Ren,Q., Tag,A., Peplow,A., Lai,H., Kupfer,C., Peterson,A., Beremand
 ,M. and Roe,B.
 TITLE Analysis of a Fusarium sporotrichioides EST database
 JOURNAL Unpublished (2001)
 COMMENT Contact: Bruce A. Roe, University of Oklahoma, broe@ou.edu
 Department of Chemistry and Biochemistry
 Advanced Center for Genome Technology, University of Oklahoma
 620 Parrington Oval, Norman, OK 73019, USA
 Tel: 405 325 4912
 Fax: 405 325 7762
 Email: broe@ou.edu
 Contact Dr. Marian Beremand regarding clone availability Included
 is the best homolog from a blastx search of Genbank nr 04-09-01
 455 5e-45 gi1805056|emb|CAA604 (X86790) N3810 [Saccharomyces
 cerevisiae]
 Seq primer: T3
 High quality sequence stop: 382.
 Location/Qualifiers
 1..519
 /organism="Fusarium sporotrichioides"
 /strain="Tri 10"
 /db_xref="taxon:5514"
 /clone="l2e03fs"
 /clone_lib="Fusarium sporotrichioides Tri 10 overexpressed
 cDNA library"
 /note="Vector: pBluescript SK-; Site_1: EcoRI; Site_2:
 XhoI; 5' end of cDNA cloned into EcoRI site of pBluescript
 XhoI; 3' end of cDNA cloned into XhoI site of pBluescript"
 BASE COUNT 137 a 166 c 116 g 100 t
 ORIGIN
 Query Match 2.0%; Score 80.4; DB 13; Length 519;
 Best Local Similarity 50.9%; Pred. No. 5.7e-10;
 Matches 217; Conservative 0; Mismatches 206; Indels 3; Gaps 1;
 QY 576 GTCCAGGTCCTCCCTATGACATCGACGGGATCAAGCCTGGGATCGTGCATTCGGTGA 635
 DB 97 GTCCAGGTCCTCCCTATGACATCGACGGGATCGTGCATTCGGTGA 635
 QY 636 GGTAACTTTTTCGAGCCCATGAGCGGCTTACGTCGAGCAGATTCTTGAACACCGCTCGG 695
 DB 157 GCGGTTTCCACAGAGCTCATCTGGCTGTATGTCGACAGAGCTCTCGAAGAGCATGGT 216
 QY 696 GACTGGCGATGTTGGTGTGGCTGACGCGGATCGCTTCAAGAAAGGCGGAG 755
 DB 217 GAGCGGATGGCGCATCGCGGATCGCGCTCTCTAAGATGCTGCCATGCCGAT 276
 QY 756 GAATTCAGGCGCCAGGACTGCTGTATTCCTGACCGAGAGCGGCTCCGCGCAAGAGC 815
 DB 277 GTCTCTCAAGCCCAAGACCTCTACACCGCTAATTTGAGCGATCGCGAAGGCGAGTTTC 336

Qy	816	ACGTGCGCGTCATGGGGCGCTGCGTGACTATCTGTTGCCCGGCAGATCCGGAAGCC	875
Db	337	GCCGA---CGTGTGAGCAGTATCACCTCCTTCATTTTCGTCGCCGACGACCGTAGGCC	393
Qy	876	GTGCTGAAGCATCTTGTGATCCGGGCATCCGCATCGTTTCCATGACCAATCAGGAAAGCC	935
Db	394	GTCAATTGCAAGAAGTGGCTCATCCCGACACTCACATCGTCTCTCACTATTACTGAGAGC	453
Qy	936	GGCTACAACATCAACGAGACGACCCTGGTGGTTCGATCTGGAGAATCGCGCAGTAAGGCC	995
Db	454	GGCTACTACTACAACAGAACACCCATCAATTGAAGGATGAGACCCCCGACATCCAGCAC	513
Qy	996	GACCTC 1001	
Db	514	GATCTC 519	

RESULT 10	BH143537	520 bp	DNA	linear	GSS 16-AUG-2001
LOCUS	BH143537				
DEFINITION	T0GE013TH cTOG Lycopersicon esculentum genomic clone cTOG27C2, DNA sequence.				
ACCESSION	BH143537				
VERSION	BH143537.1	GI:15197380			
KEYWORDS	GSS.				
SOURCE	tomato.				
ORGANISM	Lycopersicon esculentum				
	Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Asteridae; euasterids I; Solanales; Solanaceae; Solanum;				
	Lycopersicon.				
REFERENCE	1 (bases 1 to 520)				
AUTHORS	van der Hoeven,R., Sun,H., Cho,J., Utterback,T., Ronning,C. and Tanksley,S.				
TITLE	Tomato Demethylated Genomic DNA Sequences				
JOURNAL	Unpublished (2001)				

```

Contact: Cour
Clemson University Genomics Institute
Clemson University
100 Jordan Hall, Clemson, SC 29634, USA
"Email: http://www.genome.clemson.edu/orders/index.html
tomato demethylated genomic DNA
Insert Length: 1270 Std Error: 0.00
Seq primer: M13F-R
Class: Shotgun.

FEATURES             Location/Qualifiers
     source            1..520
                     /organism="Lycopersicon esculentum"
                     /cultivar="E6203"
                     /db_xref="taxon:4081"
                     /clone_lib="CTOG"
                     /clone="CTOG27C2"
                     /tissue_type="young leaves"
                     /dev_stage="12-14 weeks post harvest"
                     /lab_host="E.coli JM109"
                     /note="Vector: pBluescript SK(-); Site_1: EcoRI; Site_2:
xhoI; This library was made from short EcoRI digested
fragments of the genome of Lycopersicon esculentum ligated
into pBS (SK-). The fragments were cloned into the
methylation restrictive E.coli strain JM109 with the
purpose of enriching the library for non-methylated DNA
fragments. This procedure may enrich the pool of cloned
fragments in JM109 cells for sequences representing
expressed genes. Average insert size 1.27 kb."

75 a      181 c      186 g      78 t

BASE COUNT
ORIGIN

```

Query Match	1.9%	Score 77.6	DB 17	Length 520
Best Local Similarity	48.4%	Pred. No. 3.2e-09		
Matches 249	Conservative 0	Mismatches 259	Indels 6	Gaps 1
QY	553	CCCTTAAGTCCTCTCGCAATGTCGAGGCTCCGCCCTATGACATGCGGCGATCAAGC	612	

Db	4	CCCTGGCGCATCTGCCCGAGGAGGTGGAAATCTCCCGCCTATGACCGCAGCCGGGTCAACG	63
Qy	613	CTGGATCGTGCATTTTCGGTGTAGGTAACTTTTTCGAGCCCATGAGCGGCTTCTACCTCG	672
Db	64	CCGGCATCGTCATCTGGGCATAGGCCTTTCCACCCGCCGCCACCGCGGTGGTGACGG	123
Qy	673	AGCAGATTCTTGAACACGCTCCGAGTGGCGGATTTGGTGTGGCGTGCAGCGGCGAGTG	732
Db	124	ACTTGGGCATGACCCGCCAGCGGGGACCTAGTTGGGGCATCGTGGCGGTTTCGTGCGCA	183
Qy	733	ACCGTTCAAAGAAAAACCGAGGAATTCAAAGGCCAGGACTGCCTGTATTTCCCTGACCG	792
Db	184	GCCCCGCCACGCGACGCGCTGACGCCGAGGAGGCGCTGTACACCTTGGCGCTGCGCG	243
Qy	793	AGACGGCTCGTCCGGCAAGACGACGG-----TGGCGGTCATGGGGCGCGTGGGTGACT	846
Db	244	ACGCGGCCCGGATGGGCAGTTGCGGGAACAGTTGCGGGGTGGTGGGTGGTGGCGGG	303
Qy	847	ATCTGCTTGCCCGCGCCATCCGGAACCGTGCTGAAGCATCTTGTTGATCCCGGCCATCC	906
Db	304	TCGTGGTGGCGGAGGAATCCCCAGCGGCTGTGGAGCGCATCGCTCCGCCACGACGC	363
Qy	907	GCATCGTTTCCATGACGATCAGGAAAGCGGGCTACAACATCAACGAGACGACCGTGGCT	966
Db	364	GCATCGTCAAGCTGACCGTCCAGGAAAGGGCTACGCCAGAACCTCCACCGGCCACCC	423
Qy	967	TCGATCTGGAGATCGCGGCAGTAAAGCGCCGACCTCAAGAACCCGGAAGAACCGGCTACCG	1026
Db	424	TGGCGTGGGATGACCCGGACATCGTGACAGACCTTGGCGCATCCGGTCAAGCCCGGCGACA	483
Qy	1027	TTTTTCGGTTACGTGGTCCAGGCGCCTCGCTGCTCG	1060
Db	484	CCATCGGCATCTCGGTGTACGGCTTGGCGCTGGCG	517

RESULT 11	
AQ500091/c	
LOCUS	556 bp DNA linear GSS 29-APR-1999
DEFINITION	V34C mTn-3xHA/lacZ Insertion Library Saccharomyces cerevisiae genomic 5', DNA sequence.
ACCESSION	AQ500091
VERSION	AQ500091.1
KEYWORDS	GI:4704913
SOURCE	GSS.
ORGANISM	baker's yeast.
REFERENCE	Saccharomyces cerevisiae Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes; Saccharomycetales; Saccharomycetaceae; Saccharomyces.
AUTHORS	1 (bases 1 to 556) Ross-Macdonald, P., Roemer, T., Coelho, P.S.R., Agarwal, S., Kumar, A., desEtages, S.A., Cheung, K.-H., Sheehan, A., Symoniatidis, D., Jansen, R., Umansky, L., Heidtman, M., Nelson, K., Iwasaki, H., Kanada, D., Lugo, R., Hager, K., Miller, P., Roeder, G.S. and Snyder, M.
TITLE	Large-Scale Analysis of the Yeast Genome by Transposon Tagging and Gene Disruption
JOURNAL	Unpublished (1999)
COMMENT	Contact: Kumar A Michael Snyder, Dept. of Mol. Cell. and Dev. Biology Yale University P.O. Box 208103, New Haven, CT 06520-8103, USA Tel: 203 432 9949 Fax: 203 432 6161 Email: anuj.kumare@yale.edu te of mTn-3xHA/lacZ insertion. Seq primer: GGCTCTCTTCTTTGGAGTAGTAC Class: transposon-tagged. Location/Qualifiers 1. 556 /organism="Saccharomyces cerevisiae" /db_xref="taxon:4932" /clone_lib="mTn-3xHA/lacZ Insertion Library" /lab_host="E. coli" /note="Vector: pHS56-Sal; A yeast genomic DNA library (lacking mitochondrial DNA) was prepared in pHS56-Sal;
FEATURES	
source	

genomic DNA was size-fractionated (DNA of roughly 2-3 kb in length) prior to cloning. This library was subsequently mutagenized with a mTn-3xHA/lacZ minitransposon containing lacZ, URA3, and tet resistance."

```
BASE COUNT      146 a   129 c   126 g   155 t
ORIGIN

Query Match      1.7%; Score 69.4; DB 17; Length 556;
Best Local Similarity 52.2%; Pred. No. 4.8e-07;
Matches 179; Conservative 0; Mismatches 161; Indels 3; Gaps 1;

Qy 752 CGAGGAATTCAGGCCAGGACTGCTGTTATTCCTGACCGAGACGGCTCGTCCGGCAA 811
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 404 CGATGCCATGAAGCCCAAGATTGCCATACACCCCTGTGGAGCGGTGATCAAGGACAC 345
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Qy 812 GAGCAGCGTGGCGCTGATGCGCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 871
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 344 TAAGCGTTATATCTGTC---GGTCCATCTACTGCTTATCATGTACCTCCGATGATCCAG 288
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Qy 872 AGCGGTGCTGAAGCATCTTGTGATCGCGGCATCCGATCGCTTTCCATGACGATCACGGA 931
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 287 AGCTGTTATTGAAAGATGGCAATCCAGACACACATTTGTTCTTTGACGGGTACAGAC 228
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Qy 932 AGCGGCTACAAATCAACAGACAGACCGGTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 991
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 227 AAACGGTTACTACACAGTGAAGCAACAACCTCCTTAATGACAGATGCTCCCGAGATTAT 168
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Qy 992 GGCGGACCTCAAGAACCCGGAAGCCGCTCTACCGTTTTTCGGTTACGTGGTFCGAGGCCCT 1051
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 167 CAATGATTGTAACCAACCCAGAAAGCCAGATCTCTGTATGGGTACCTATATGAAGCCCT 108
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Qy 1052 CGCTGCTGTTGGATGCCGCTGTTGAAGGCAATTTACGGTCAATG 1094
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 107 GTTGTTGGTTTACAAGAGAGCTCTTACCCCATTCCTATTAGG 65
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
```

```
RESULT 12
BG608411
LOCUS      BG608411          506 bp      mRNA      linear      EST 17-APR-2001
DEFINITION 296895 MARC 1PIG Sus scrofa cDNA 5', mRNA sequence.
ACCESSION  BG608411
VERSION    BG608411.1  GI:13658389
KEYWORDS   EST.
SOURCE     pig.
ORGANISM   Sus scrofa
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Cetartiodactyla; Suina; Suidae; Sus.
REFERENCE  1 (bases 1 to 506)
AUTHORS   Fahrenkrug,S.C., Freking,B.A., Rohrer,G.A., Smith,T.P.L., Casas,E.,
            Stone,R.T., Heaton,M.P., Grosse,W.M., Bennett,G.A., Laegreid,W.W.
            and Keele,J.W.
TITLE      Design and use of two pooled tissue normalized cDNA libraries for
            EST discovery in swine
JOURNAL    Unpublished (2000)
COMMENT    Contact: Smith TPL
            USDA, ARS, US Meat Animal Research Center
            PO Box 166, Clay Center, NE 68933-0166, USA
            Tel: 402 762 4366
            Fax: 402 762 4390
            Email: smithemail.marc.usda.gov
            Single pass sequencing. Bases called and alt_trimmed with phred
            v0.980904.e. Vector identified by cross_match with the -minscore 18
            and -minmatch 12 options.
PCR PRIMERS
FORWARD: AGGAACAGCTATGACCAT
BACKWARD: GTTTCCTCCAGTCAGCAGG
Plate: 1 row: A column: 22
Seq primer: ATTAGGTGACACTATAG.
Location/Qualifiers
            1..506
            /organism="Sus scrofa"
            /db_xref="taxon:9823"
            /clone_lib="MARC 1PIG"
```

```
/tissue_type="pooled"
/lab_host="DHI0B"
/note="vector: pCMV SPORT6; Site_1: NotI; Site_2: SalI;
Library made from pooled tissue from day 11, 13, 15, 20,
and 30 embryos."
BASE COUNT      130 a   118 c   138 g   120 t
ORIGIN

Query Match      1.3%; Score 54.2; DB 12; Length 506;
Best Local Similarity 49.8%; Pred. No. 0.0048;
Matches 137; Conservative 0; Mismatches 138; Indels 0; Gaps 0;

Qy 1241 TTGCGCGGAATGCCCAAGAGCTCAACGGCGGCAGTGGGTGATGATGACGACCTGCCGCT 1300
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 2 TGCACCGACATACTGTTGGAAATACGACGACATCTGGGGTGAATGATCCCTCGCGCAT 61
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Qy 1301 GGTGGCGGAGGATTTCCATCAGTGGGTGCTGGAAGACACAGTTTTCGCGATGCCGCTCCGCC 1360
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 62 TAGTTCGGAACCGTTTATCCAGTGGGTGGTGGGAAGATAACTTCATCCCTGGGCGCTCTGC 121
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Qy 1361 GCTTGAAGAAAGCCGCGCTGCAGATGTCGGGGAGCGTGACGGACTGGGAGTACGTCACAGT 1420
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 122 CTGGGAAGTCGCAGATCTACAAATGGTGAATGATGTCTGCCATGGGAAGAGATGAAACT 181
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Qy 1421 CCGAATCTCAATGACGGGATGTGATGCTGCTGCTCCAGGCATTTCTGGTCGCGCTATGA 1480
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 182 GCGGATCTTAATGGCAGCCACTCTTTTCTCGCTTATCTGGGTTACCTCTCAGGATTCCG 241
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Qy 1481 GAATGTGGATGACGCCATTTGAAGACAGCACTCC 1515
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 242 CCATATCAGTGTGATGAGGATCGCGCATTC 276
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

RESULT 13
AZ933800
LOCUS      AZ933800          788 bp      DNA      linear      GSS 24-APR-2001
DEFINITION BJ_Ba0001G08f B. japonicum BAC library Bradyrhizobium japonicum
            genomic, DNA sequence.
ACCESSION  AZ933800
VERSION    AZ933800.1  GI:13775860
KEYWORDS   GSS.
SOURCE     Bradyrhizobium japonicum.
ORGANISM   Bradyrhizobium japonicum
            Bacteria; Proteobacteria; Alphaproteobacteria; Rhizobiales;
            Bradyrhizobiaceae; Bradyrhizobium.
REFERENCE  1 (bases 1 to 788)
AUTHORS   Tomkins,J.P., Wood,T.C., Stacey,M.G., Loh,J.T., Judd,A., Golcochea
            ,J.L., Stacey,G., Sadowsky,M.J. and Wing,R.A.
            A marker-dense, sequence-ready map of the Bradyrhizobium japonicum
            genome
            Genome Res. 11 (8), 1434-1440 (2001)
JOURNAL    21376150
MEDLINE    Contact: Wing RA
COMMENT    Clemson University Genomics Institute
            Clemson University
            100 Jordan Hall, Clemson, SC 29634, USA
            Tel: 864 656 7288
            Fax: 864 656 4293
            Email: rwing@clemson.edu
            Class: BAC ends
            High quality sequence stop: 725.
            Location/Qualifiers
                1..788
                /organism="Bradyrhizobium japonicum"
                /strain="USD110"
                /db_xref="taxon:375"
                /clone_lib="B. japonicum BAC library"
                /lab_host="E. coli"
                /note="vector: pindigo536; Site_1: HindIII"
BASE COUNT      128 a   271 c   258 g   131 t
ORIGIN

Query Match      1.2%; Score 51.4; DB 17; Length 788;
```


	Best Local Similarity	52.5%; Matches 137;	Pred. No. 0.034; Conservative 0;	Mismatches 121;	Indels 3;	Gaps 1;
Qy	1422	CGAATGCTCAATGCAGGGCATGT	CATCCTCTGCTTCCAGGCCATTCTGGTCGGCTATGAG	1481		
Dd	1	CGGCTGCTCAACGCCAGCCAFT	CGCGCTCGCCTATCTCGGCTATCTCGCGGCTACGAG	60		
Qy	1482	AATGTGGATGACGCCCAATTGAAGACAGCGAACTCCTTTGGCAATCTGAAGA	ACTATCTCAAC	1541		
Dd	61	ACCATCGCCGACACCATGCAAGATCCGC	ACTTC---CGCGCTCTCGCGCGCGCAGGTGATG	117		
Qy	1542	AAGGATGTCATCCGACCCTTGAGGGCCCTTCAGGCATGACGCTCGAAGGCTATCGGGAC	1601			
Dd	118	GAGAGCGCCGGTAGCGCTGAAGATCCGGCGGACGACCGACCTTGC	CGGCTATCTCGCGCC	177		
Qy	1602	AGCGTCATCAGCCGTTTTCTCCAAACAAGCGCATGTTCGGAACGAGCGCTCCGGATTGCTAGC	1661			
Dd	178	TCGCTGCTCAAGCGCTTCGCCAATCCGGCGCTGCATCATCGACCTGGCAGATCGCGATG	237			
Qy	1662	GATGGCTGTTTCCAAAGGTTTCAG	1682			
Dd	238	GACGGCTCGCAAAGGCTGCGG	258			

RESULT 14	CNS017SY	1101 bp	DNA	linear	GSS 26-JUL-1999
LOCUS	CNS017SY/c				
DEFINITION	Drosophila melanogaster genome survey sequence SP6 end of BAC BACN37L08 of DrosBAC library from Drosophila melanogaster (fruit fly), genomic survey sequence.				
ACCESSION	AL108460				
VERSION	AL108460.1	GI:5628764			
KEYWORDS	GSS.				
SOURCE	Drosophila melanogaster.				
ORGANISM	Drosophila melanogaster. Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.				
REFERENCE	1 (bases 1 to 1101)				
AUTHORS	Genoscope.				
TITLE	Direct Submission				
JOURNAL	Submitted (23-JUL-1999); Genoscope - Centre National de Sequencage : BP 191 91006 EVRY cedex - FRANCE (E-mail : seqref@genoscope.cns.fr - Web : www.genoscope.cns.fr)				
COMMENT	Determination of this BAC-end sequence was carried out as part of a collaboration with the European Drosophila Genome Project (EDGP) - http://www.edgp.ebi.ac.uk -. This Drosophila melanogaster BAC library (Dros BAC) was made by Alain Billaud at CEPH (Centre d'Etude du Polymorphisme Humain) with funding provided by a MRC project grant. The DNA was prepared from embryos by Alain Bucheton and Genevieve Payan. It has been constructed in the vector pBelobAC11.				

FEATURES	source	Location/Qualifiers
pbe08A011.	1..1101	/organism="Drosophila melanogaster"
	/db_xref="taxon:7227"	
	/clone_lib="BACN37L08"	
	/clone_lib="DrosBAC"	
	/plasmid="pBelobAC11"	
	/note="end : SP6"	
BASE COUNT	254 a 176 c 160 g 152 t	359 others.
ORIGIN		
Query Match	1.2%	Score 49.2; DB 17; Length 1101;
Best Local Similarity	16.1%;	Pred. No. 0.16;
Matches	62; Conservative	166; Mismatches 158; Indels 0; Gaps 0;
QY	2318	TGCGGGCCACC CGGTCTCTGFCGCCGTTGGCTCTCTCATAGGTTTCTGGCCA 2377
Dd	1049	BSSBGSGSSBSBSSTSSSSSSSTSSSSSTSSSBFBTBTSSSSSTTTTTTTTTTTT 990
QY	2378	GCACGTCCGATGATCTCCGGATCAGGTCGGCCGACGGCCACGCGGATTCTTGTCG 2437

Db	989	BTBBSBSBSSSSSTSSBBSBTSKBSBTSBSSBSBSSSSASBSBSSSSSSSSSSBSSSTTTBTSBBSB	930		
Qy	2438	TTGCGTGCGGGTGATCCCGAGAAATACGATAGCATCCGGCTCGTTTCCCGCTGGCGGCGC	2497		
Db	929	STSSSSGSSSSBBSBTSBTTTBTBKSTSTSSSTSBTTTBSBSSSSSSBS	870		
Qy	2498	GATTGTTGCGCTTTCGGCCCGGTCCCATCTCTCGCGGCGAGCCAAATGCCCGGTGAA	2557		
Db	869	SSBTSSTTBSSTSSSYSSSBTCSKSTBTSBSTBTSBMSKBSSTTSBTSBSSBGC	810		
Qy	2558	CGCGCTCACAAAATCATCTCTTCGGGTGAAGTCGCGGCTGGCGCGCATCGGCAC	2617		
Db	809	GSGSTSGSSCSBSCGKSTSSGSTSGTTCGBCSSGCGSCCCYCSCCCTCTCGTSTCC	750		
Qy	2618	GGCGGATACGAACAGTCCGTCATGAGTTCTCAAGCGCGCGTATTATTCGCGCATAGG	2677		
Db	749	SCSGGTGSSBKCYSTSSASSTSSSGTSSSSGTTGTCGCGGTTTGTGTCGCCCTG	690		
Qy	2678	CCTTGCCCATTTCCGCGGCATATC	2703		
Db	689	GBYCCYBCCMTMCCGSSCCSCSYC	664		
RESULT 15	AL539916	861 bp	mRNA	linear	EST 16-FEB-2001
LOCUS	AL539916	MTI_FL013_FBRn1	Homo sapiens	CDNA clone	CSODF031YMW10 5
DEFINITION	AL539916	primo, mRNA sequence.			
ACCESSION	AL539916				
VERSION	AL539916.1	GI:12869578			
KEYWORDS	EST.				
SOURCE	human.				
ORGANISM	Homo sapiens				
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.				
AUTHORS	1. (bases 1 to 861)				
TITLE	Li, W.B., Gruber, C., Jessee, J., and Polayes, D.				
JOURNAL	Full-length cDNA libraries and normalization				
COMMENT	Unpublished (2001)				
Genoscope - Centre National de Sequencage					
BP 191 91006 EVRY cedex - France					
Email: seqref@genoscope.cns.fr, Web : www.genoscope.cns.fr.					
Location/Qualifiers					
source	1. .861				

```

FEATURES
source
1. 861
Location/Qualifiers
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="CS0DF031YM10"
/clone_lib="LTI_FLO13_FBRn1"
/dev_stage="pooled tissue from post conception fetuses (20
week, 24 week and 26 week)"
/lab_host="DH108"
note="Organ: Fetal brain; Vector: pCMVSPORT 6; 1st strand
cDNA was primed with a NotI-oligo(dT) primer. Five prime
end enriched, double-stranded cDNA was digested with Not I
and cloned into the Not I and Eco RV sites of the
pCMVSPORT 6 vector. Library was constructed by Life
Technologies. Contact : Feng Liang Life Technologies, a
division of Invitrogen 9800 Medical Center Drive Rockville
, Maryland 20850, USA Fax : (1) 301 610 8371 Email :
fliang@lifetech.com URL :
http://fulllength.invitrogen.com"
BASE COUNT      188 a      282 c      243 g      147 t      1 others
ORIGIN

Query Match      1.2%; Score 48; DB 9; Length 861;
Best Local Similarity 49.6%; Pred. No. 0.28;
Matches 123; Conservative 0; Mismatches 125; Indels 0; Gaps 0;

QY      761 CAAGGCCCGGACTGCTGTATTCCCTGACCGAGCGGTCGCTCGCGGCAAGACGACGGT 820
      || | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db      319 CATTGTTCATGACTCTGTGAGACGGGGGTTCACCCACAGCGGTGCCCATCTACGAGGGCTACGC 378

```


GenCore version 5.1.4.p5.4578
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: May 8, 2003, 16:34:43 ; Search time 40 Seconds
(without alignments)
1615.664 Million cell updates/sec

Title: US-09-926-163B-2

Perfect score: 485

Sequence: 1 MITRETLKSLPANVQAPPD.....KVIVLRKIIRKGVKAIPA 485

Scoring table: OLIGO

Gapop 60.0 , Gapext 60.0

Searched: 908470 seqs, 133250620 residues

Word size : 0

Total number of hits satisfying chosen parameters: 908470

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Database : A_Geneseq_101002.*

```

1: /SID52/gcgdata/geneseq/geneseq-emb1/AA1980.DAT.*
2: /SID52/gcgdata/geneseq/geneseq-emb1/AA1981.DAT.*
3: /SID52/gcgdata/geneseq/geneseq-emb1/AA1982.DAT.*
4: /SID52/gcgdata/geneseq/geneseq-emb1/AA1983.DAT.*
5: /SID52/gcgdata/geneseq/geneseq-emb1/AA1984.DAT.*
6: /SID52/gcgdata/geneseq/geneseq-emb1/AA1985.DAT.*
7: /SID52/gcgdata/geneseq/geneseq-emb1/AA1986.DAT.*
8: /SID52/gcgdata/geneseq/geneseq-emb1/AA1987.DAT.*
9: /SID52/gcgdata/geneseq/geneseq-emb1/AA1988.DAT.*
10: /SID52/gcgdata/geneseq/geneseq-emb1/AA1989.DAT.*
11: /SID52/gcgdata/geneseq/geneseq-emb1/AA1990.DAT.*
12: /SID52/gcgdata/geneseq/geneseq-emb1/AA1991.DAT.*
13: /SID52/gcgdata/geneseq/geneseq-emb1/AA1992.DAT.*
14: /SID52/gcgdata/geneseq/geneseq-emb1/AA1993.DAT.*
15: /SID52/gcgdata/geneseq/geneseq-emb1/AA1994.DAT.*
16: /SID52/gcgdata/geneseq/geneseq-emb1/AA1995.DAT.*
17: /SID52/gcgdata/geneseq/geneseq-emb1/AA1996.DAT.*
18: /SID52/gcgdata/geneseq/geneseq-emb1/AA1997.DAT.*
19: /SID52/gcgdata/geneseq/geneseq-emb1/AA1998.DAT.*
20: /SID52/gcgdata/geneseq/geneseq-emb1/AA1999.DAT.*
21: /SID52/gcgdata/geneseq/geneseq-emb1/AA2000.DAT.*
22: /SID52/gcgdata/geneseq/geneseq-emb1/AA2001.DAT.*
23: /SID52/gcgdata/geneseq/geneseq-emb1/AA2002.DAT.*

```

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	485	100.0	485	21	AAB231172
2	52	10.7	485	20	AAV49913
3	10	2.1	10	21	AAB231173
4	9	1.9	503	22	AAG89872
5	9	1.9	503	22	AAV79410
6	8	1.6	8	20	AAV49914
7	8	1.6	8	20	AAV49915
8	8	1.6	103	23	ABP33539
9	8	1.6	163	21	AAV99387
10	8	1.6	163	22	AAU12423

Gluconobacter oxyd
 Gluconobacter subo
 Gluconobacter oxyd
 C glutamicum prote
 Corynebacterium gl
 Gluconobacter subo
 Gluconobacter subo
 Human reductase-11
 Human PRO1273 (UNQ
 Human PRO1273 poly

11	8	1.6	163	22	AAB66136	Protein of the inv
12	8	1.6	163	22	AAB36671	Human secretory pr
13	8	1.6	163	23	AB955524	Human angiogenesis
14	8	1.6	163	23	AB84918	Human PRO1273 prot
15	8	1.6	163	23	AB84918	Human PRO protein,
16	8	1.6	164	20	AAU83683	Human secreted pro
17	8	1.6	172	22	AB95653	Human testicular a
18	8	1.6	172	22	AAU94192	Human reproductiv
19	8	1.6	261	22	AB860855	Drosophila melanog
20	8	1.6	580	23	AB855021	Lactococcus lactis
21	8	1.6	1006	23	AB893763	Herbicidally activ
22	7	1.4	43	22	AB843361	Peptide #10867 enc
23	7	1.4	43	22	AAU64271	Human brain expres
24	7	1.4	69	21	AAU52675	Arabidopsis thalia
25	7	1.4	72	22	AAU59450	Propionibacterium
26	7	1.4	73	22	AB843223	Peptide #10729 enc
27	7	1.4	73	22	AB826304	Protein #8303 enco
28	7	1.4	73	22	AAU64131	Human brain expres
29	7	1.4	73	22	AAU76955	Human bone marrow
30	7	1.4	73	22	AAU21040	Peptide #7474 enco
31	7	1.4	73	22	AAU37063	Peptide #11100 enc
32	7	1.4	73	23	AB846062	Human peptide enco
33	7	1.4	98	22	AAU54359	Propionibacterium
34	7	1.4	98	22	AB843453	Peptide #10959 enc
35	7	1.4	98	22	AB826421	Protein #8420 enco
36	7	1.4	98	22	AAU64382	Human brain expres
37	7	1.4	98	22	AAU77204	Human bone marrow
38	7	1.4	98	22	AAU21137	Peptide #7571 enco
39	7	1.4	98	22	AAU37341	Peptide #11378 enc
40	7	1.4	98	23	AB846215	Human peptide enco
41	7	1.4	106	23	ABP41635	Human ovarian anti
42	7	1.4	132	22	AAU22384	Human cardiovascular
43	7	1.4	136	21	AAU01518	Human secreted pro
44	7	1.4	142	22	AAU80178	Human protein SEQ
45	7	1.4	143	22	AAU90408	C glutamicum prote

ALIGNMENTS

RESULT 1
 AAB231172
 ID AAB231172 standard; Protein; 485 AA.
 AC AAB231172;
 XX
 XX
 DT 29-JAN-2001 (first entry)
 DE
 DE Gluconobacter oxydans D-sorbitol dehydrogenase (SLDH).
 XX
 DE D-sorbitol dehydrogenase; SLDH; L-sorbose production; NADP+ dependent;
 KW sorbitol; mannitol; arabitol; oxidation; 2-keto-L-gluconic acid;
 KW L-ascorbic acid biosynthesis; vitamin C.
 XX
 OS Gluconobacter oxydans.
 OS
 PN WO200055329-A1.
 XX
 PD 21-SEP-2000.
 XX
 PF 16-MAR-2000; 2000WO-JF01608.
 XX
 PR 17-MAR-1999; 99JP-0072810.
 PR 06-AUG-1999; 99JP-0224679.
 XX
 PA (FUJI) FUJISAWA PHARM CO LTD.
 XX
 PI Shibata T, Ichikawa C, Matsuura M, Noguchi Y, Saito Y;
 PI Yamashita M, Takata Y;
 XX
 DR WPI; 2000-587530/55.
 DR N-PSDB; AAA97430.
 XX

PT Gluconobacter oxydans derived D-sorbitol dehydrogenase obtainable by
PT culturing its gene-transformed host cells, useful for producing
PT L-sorbose or 2-keto-L-gluconic acid as an intermediate in L-ascorbic
PT acid production
PS Claim 5; Page 58-60; 72pp; Japanese.
XX
CC The invention relates to Gluconobacter oxydans D-sorbitol dehydrogenase
CC (SLDH; AAB23172) and to the gene encoding it (AAA97430). SLDH has a
CC molecular weight of about 54 kDa and catalyses the conversion of
CC D-sorbitol into L-sorbose using NADP+ as a co-enzyme. SLDH specifically
CC catalyses the oxidation of sorbitol, mannitol and arabinol, but does not
CC act on xylitol, ribitol, inositol and glycerol. The invention also
CC encompasses expression vectors and host cells comprising the
CC Gluconobacter oxydans SLDH gene, and the recombinant production of SLDH.
CC The invention further relates to a method for preparing L-sorbose by
CC contacting the recombinant SLDH with D-sorbitol; a process for producing
CC 2-keto-L-gluconic acid by contacting recombinantly produced sorbose
CC dehydrogenase and/or sorbose dehydrogenase with L-sorbose; and a
CC process for preparing L-ascorbic acid or its alkaline earth metals salts
CC by conversion from 2-keto-L-gluconic acid. SLDH is useful for producing
CC L-sorbose or 2-keto-L-gluconic acid as an intermediate in L-ascorbic
CC acid production. The present sequence represents the Gluconobacter
CC oxydans SLDH protein.
XX
SQ Sequence 485 AA;

Query Match 100.0%; Score 485; DB 21; Length 485;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 485; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 MITRETLSPANVQAPPYDIDIKPGIVHFGVGNFFRAHEAFYVEQILEHAPDWAIVGV 60
DB 1 MITRETLSPANVQAPPYDIDIKPGIVHFGVGNFFRAHEAFYVEQILEHAPDWAIVGV 60
QY 61 GLUTGSDRSKKKAEFEKAQDCLYSLTETAPSGKSTVRVMGALRDYLLAPADPEAVLKHLVD 120
DB 61 GLUTGSDRSKKKAEFEKAQDCLYSLTETAPSGKSTVRVMGALRDYLLAPADPEAVLKHLVD 120
QY 121 PAIRIVSMTITEGGYNINETTGAFDLENAVKADLNKPEKSTVFYVVEALRRRWDAGG 180
DB 121 PAIRIVSMTITEGGYNINETTGAFDLENAVKADLNKPEKSTVFYVVEALRRRWDAGG 180
QY 181 KFTVMSCDNLRHNGNARKAFGLGAKARDEPELAKWIEENATFPNGMVDRTPTVSA 240
DB 181 KFTVMSCDNLRHNGNARKAFGLGAKARDEPELAKWIEENATFPNGMVDRTPTVSA 240
QY 241 KKLNAASGLDDDLPLVAEDFHQWLEDFADGRPPLEKAGVQMGVDVTDWEYVKIRMLNA 300
DB 241 KKLNAASGLDDDLPLVAEDFHQWLEDFADGRPPLEKAGVQMGVDVTDWEYVKIRMLNA 300
QY 301 GHVMLCFPGILVGENVDADAIEDSELLGNLKNLKNVDIPTLKAPSGMTLEGYRDSVISR 360
DB 301 GHVMLCFPGILVGENVDADAIEDSELLGNLKNLKNVDIPTLKAPSGMTLEGYRDSVISR 360
QY 361 FSNKAMSQDTLRASDGGSKVQVFTWTRRAIEDKRLSRIAFGSIASYLEMLRGRDEKG 420
DB 361 FSNKAMSQDTLRASDGGSKVQVFTWTRRAIEDKRLSRIAFGSIASYLEMLRGRDEKG 420
QY 421 GYSESSEPTYGDAENKLAADDFESSLKLPAFGDGRDLDTSELDOKVIVLRKIIREKGVK 480
DB 421 GYSESSEPTYGDAENKLAADDFESSLKLPAFGDGRDLDTSELDOKVIVLRKIIREKGVK 480
QY 481 AAIAPA 485
DB 481 AAIAPA 485

RESULT 2
AAV49913
ID AAV49913 standard; Protein: 485 AA.
XX
AC AAV49913;

XX 27-JAN-2000 (first entry)
DT Gluconobacter suboxydans L-sorbose reductase protein sequence.
XX
DE Gluconobacter suboxydans; L-sorbose reductase; genetic engineering;
XX mutant.
KW Gluconobacter suboxydans.
XX
OS Gluconobacter suboxydans.
XX
PN AU9920390-A.
XX
PD 23-SEP-1999.
XX
PF 11-MAR-1999; 99AU-0020390.
XX
PR 13-MAR-1998; 98EP-0104546.
XX
PA (HOFF) HOFFMANN LA ROCHE & CO AG F.
XX
PI Hoshino T, Tazoe M, Shinjoh M, Kon T;
XX
DR WPI; 1999-579275/49.
DR N-PSDB; AAZ35672.
XX
PT Set of genetically modified mutants not containing L sorbose reductase
XX
PS Claim 8; Page 18-21; 33pp; English.
XX
CC The present invention describes a genetically engineered
CC microorganism derived from a microorganism belonging to the genus
CC Gluconobacter or Acetobacter which is characterised in that the
CC biological activity for reducing L-sorbose is substantially
CC nullified by gene recombination gene. The present sequence represents
CC Gluconobacter suboxydans L-sorbose reductase.
CC N.B. This patent is equivalent to the basic NO9901197 in week 199949.
XX
SQ Sequence 485 AA;
Query Match 10.7%; Score 52; DB 20; Length 485;
Best Local Similarity 100.0%; Pred. No. 3 6e-43;
Matches 52; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 186 MSCDNLRHNGNARKAFGLGAKARDEPELAKWIEENATFPNGMVDRTPTVSA 237
DB 186 MSCDNLRHNGNARKAFGLGAKARDEPELAKWIEENATFPNGMVDRTPTVSA 237
RESULT 3
AAB23173
ID AAB23173 standard; peptide; 10 AA.
XX
AC AAB23173;
XX
DT 29-JAN-2001 (first entry)
XX
DE Gluconobacter oxydans SLDH N-terminal sequence.
XX
KW D-sorbitol dehydrogenase; SLDH; L-sorbose production; NADP+ dependent;
KW sorbitol; mannitol; arabinol; oxidation; 2-keto-L-gluconic acid;
KW L-ascorbic acid biosynthesis; vitamin C; N-terminus.
XX
OS Gluconobacter oxydans.
XX
PN WO200055329-A1.
XX
PD 21-SEP-2000.
XX
PF 16-MAR-2000; 2000WO-JP01608.
XX
PR 17-MAR-1999; 99JP-0072810.
PR 06-AUG-1999; 99JP-0224679.
XX

PA (FUJI) FUJISAWA PHARM CO LTD.
 XX Shibata T, Ichikawa C, Matsuura M, Noguchi Y, Saito Y;
 PI Yamashita M, Takata Y;
 XX WPI: 2000-587530/55.
 XX Gluconobacter oxydans derived D-sorbitol dehydrogenase obtainable by
 PT culturing its gene-transformed host cells, useful for producing
 PT L-sorbose or 2-keto-L-gluconic acid as an intermediate in L-ascorbic
 PT acid production
 XX
 PS Example 2; Page 28; 72pp; Japanese.
 XX
 CC The invention relates to Gluconobacter oxydans D-sorbitol dehydrogenase
 CC (SLDH; AAB331172) and to the gene encoding it (AAA97430). SLDH has a
 CC molecular weight of about 54 kDa and catalyses the conversion of
 CC D-sorbitol into L-sorbose using NADP+ as a co-enzyme. SLDH specifically
 CC catalyses the oxidation of sorbitol, mannitol and arabitol, but does not
 CC act on xylitol, ribitol, inositol and glycerol. The invention also
 CC encompasses expression vectors and host cells comprising the
 CC Gluconobacter oxydans SLDH gene, and the recombinant production of SLDH.
 CC The invention further relates to a method for preparing L-sorbose by
 CC contacting the recombinant SLDH with D-sorbitol; a process for producing
 CC 2-keto-L-gluconic acid by contacting recombinantly produced sorbose
 CC dehydrogenase and/or sorbosone dehydrogenase with L-sorbose; and a
 CC process for preparing L-ascorbic acid or its alkaline earth metals salts
 CC by conversion from 2-keto-L- gluconic acid. SLDH is useful for producing
 CC L-sorbose or 2-keto-L- gluconic acid as an intermediate in L-ascorbic
 CC acid production. The present sequence represents the N-terminal sequence
 CC of the Gluconobacter oxydans SLDH protein obtained in an exemplification
 CC of the invention.
 XX
 SQ Sequence 10' AA;
 Query Match 2.1%; Score 10; DB 21; Length 10;
 Best Local Similarity 100.0%; Pred. No. 0.015;
 Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 1 MITRETLKSL 10
 DB 1 MITRETLKSL 10
 |||||
 RESULT 4
 AAG89872
 ID AAG89872 standard; Protein: 503 AA.
 AC AAG89872;
 XX
 DT 26-SEP-2001 (first entry)
 XX
 DE C glutamicum protein fragment SEQ ID NO: 3626.
 XX
 KW Coryneform bacterium; amino acid synthesis; vitamin; saccharide;
 KW organic acid synthesis.
 XX
 OS Corynebacterium glutamicum.
 XX
 PN EP1108790-A2.
 XX
 PD 20-JUN-2001.
 XX
 PF 18-DEC-2000; 2000EP-0127688.
 XX
 PR 16-DEC-1999; 99JP-0377484.
 PR 07-APR-2000; 2000JP-0159162.
 PR 03-AUG-2000; 2000JP-0280988.
 XX
 PA (KYOW) KYOWA HAKKO KOGYO KK.
 XX Nakagawa S, Mizoguchi H, Ando S, Hayashi M, Ochiai K, Yokoi H;
 PI Tateishi N, Senoh A, Ikeda M, Ozaki A;
 WPI: 2001-376931/40.
 N-PSDB; AAH65091.
 Novel polynucleotides derived from Coryneform bacteria, for identifying
 mutation point of a gene, measuring expression of a gene, analysing
 expression profile or pattern of a gene and identifying homologous gene
 Claim 17; SEQ ID NO: 3626; 246pp + Sequence Listing; English.
 The present invention provides a number of nucleotide and protein
 sequences from the Coryneform bacterium Corynebacterium glutamicum. These
 are useful for identifying the mutation point of a gene derived from a
 mutant of coryneform bacterium, measuring expression amount and
 analysing the expression profile or expression pattern of a gene derived
 from Coryneform bacterium, and identifying a homologue of a gene derived
 from coryneform bacterium. Coryneform bacteria are useful for producing
 amino acids, nucleic acids, vitamins, saccharides and organic acids,
 particularly L-lysine. The present sequence is a protein described
 in the exemplification of the invention.
 Note: The sequence data for this patent did not form part of the printed
 specification, but was obtained in electronic format directly from the
 European Patent Office.
 Query Match 1.9%; Score 9; DB 22; Length 503;
 Best Local Similarity 100.0%; Pred. No. 4.5;
 Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 26 PGIVHFGVG 34
 DB 33 PGIVHFGVG 41
 |||||
 RESULT 5
 AAB79410
 ID AAB79410 standard; Protein: 503 AA.
 AC AAB79410;
 XX
 DT 30-APR-2001 (first entry)
 XX
 DE Corynebacterium glutamicum SMP protein sequence SEQ ID NO:336.
 XX
 KW Corynebacterium glutamicum; carbon metabolism and energy production;
 KW SMP protein; sugar metabolism and oxidative phosphorylation protein;
 KW fine chemical production; organic acid; proteinogenic amino acid;
 KW nonproteinogenic amino acid; purine base; pyrimidine base; nucleoside;
 KW nucleotide; lipid; saturated fatty acid; unsaturated fatty acid; diol;
 KW carboxylate; aromatic compound; vitamin; cofactor; polypeptide; enzyme;
 KW diagnosis; Corynebacterium diphtheriae; evolutionary study.
 XX
 OS Corynebacterium glutamicum.
 XX
 PN WO200100844-A2.
 XX
 PD 04-JAN-2001.
 XX
 PF 23-JUN-2000; 2000WO-IB00943.
 XX
 PR 25-JUN-1999; 99US-0141031.
 PR 08-JUL-1999; 99DE-1031412.
 PR 08-JUL-1999; 99DE-1031413.
 PR 08-JUL-1999; 99DE-1031419.
 PR 08-JUL-1999; 99DE-1031420.
 PR 08-JUL-1999; 99DE-1031424.
 PR 08-JUL-1999; 99DE-1031428.
 PR 08-JUL-1999; 99DE-1031431.
 PR 08-JUL-1999; 99DE-1031433.
 PR 08-JUL-1999; 99DE-1031434.
 PR 08-JUL-1999; 99DE-1031510.

PR 08-JUL-1999; 99DE-1031562.
 PR 08-JUL-1999; 99DE-1031634.
 PR 09-JUL-1999; 99DE-1032180.
 PR 09-JUL-1999; 99DE-1032227.
 PR 09-JUL-1999; 99DE-1032230.
 PR 09-JUL-1999; 99DE-1032230.
 PR 09-JUL-1999; 99DE-1032230.
 PR 14-JUL-1999; 99DE-1032924.
 PR 14-JUL-1999; 99DE-1032973.
 PR 14-JUL-1999; 99DE-1033005.
 PR 14-JUL-1999; 99DE-1040765.
 PR 27-AUG-1999; 99DE-1040765.
 PR 31-AUG-1999; 99DE-1040765.
 PR 03-SEP-1999; 99DE-1042076.
 PR 03-SEP-1999; 99DE-1042079.
 PR 03-SEP-1999; 99DE-1042086.
 PR 03-SEP-1999; 99DE-1042087.
 PR 03-SEP-1999; 99DE-1042088.
 PR 03-SEP-1999; 99DE-1042095.
 PR 03-SEP-1999; 99DE-1042123.
 PR 03-SEP-1999; 99DE-1042125.
 XX
 PA (BADI) BASF AG.
 XX
 PI Pompejus M, Kröger B, Schroeder H, Zelder O, Haberhauer G;
 DR
 DR WPI; 2001-061975/07.
 DR N-PSDB; AAF71527.
 XX
 XX New isolated Corynebacterium glutamicum nucleic acid encoding a sugar
 PT metabolism and oxidative phosphorylation protein for production or
 PT modulation of production of fine chemicals e.g. amino acids,
 PT carbohydrates or enzymes -
 XX
 XX Claim 20; Page 615-616; 1246pp; English.
 XX
 CC AAF71360 to AAF71750 encode the Corynebacterium glutamicum sugar
 CC metabolism and oxidative phosphorylation (SMP) proteins given in
 CC AAB79243 to AAB 79633 which are involved in carbon metabolism and
 CC energy production. The C. glutamicum SMP gene can be used in vectors
 CC (II) for expression in host cells and production or modulation of
 CC production of fine chemicals, such as, an organic acid, a proteinogenic
 CC or nonproteinogenic amino acid (preferred), a purine or pyrimidine base,
 CC a nucleoside, a nucleotide, a lipid, a saturated or unsaturated fatty
 CC acid, a diol, a carbohydrate, an aromatic compound, a vitamin, a
 CC cofactor, a polyketide, or an enzyme. The presence of (I) or SMP proteins
 CC (III) encoded by them are used for diagnosing the presence or activity of
 CC Corynebacterium diphtheriae in a subject. (I), (II), (III) or host cells
 CC containing them are used to map genomes of organisms related to
 CC C. glutamicum, identify and localise C. glutamicum sequences of interest,
 CC in evolutionary studies, in determining SMP protein regions required
 CC for function, in modulating SMP protein activity, in modulating the
 CC metabolism of sugars, and in modulating high-energy molecule production
 CC in a cell (i.e. ATP, NADPH).
 XX
 XX Sequence 503 AA;
 SQ
 Query Match 1.9%; Score 9; DB 22; Length 503;
 Best Local Similarity 100.0%; Pred. No. 4.5;
 Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 26 PGIVHFGVG 34
 DB 33 PGIVHFGVG 41
 RESULT 6
 AAY49914
 ID AAY49914 standard; peptide: 8 AA.
 XX
 AC AAY49914;
 XX
 DT 27-JAN-2000 (first entry)
 XX
 DE Gluconobacter suboxydans L-sorbose reductase peptide SEQ ID NO:3.

XX Gluconobacter suboxydans; L-sorbose reductase; genetic engineering;
 KW mutant.
 XX
 OS Gluconobacter suboxydans.
 XX
 PN AU9920390-A.
 PD 23-SEP-1999.
 XX
 PF 11-MAR-1999; 99AU-0020390.
 XX
 PR 13-MAR-1998; 98EP-0104546.
 XX
 PA (HOFF) HOFFMANN LA ROCHE & CO AG F.
 XX
 PI Hoshino T, Tazoe M, Shinjoh M, Kon T;
 XX
 DR WPI; 1999-579276/49.
 XX
 PT Set of genetically modified mutants not containing L sorbose reductase
 XX
 PS Example 3; Fig 2; 33pp; English.
 XX
 CC The present invention describes a genetically engineered
 CC microorganism derived from a microorganism belonging to the genus
 CC Gluconobacter or Acetobacter which is characterised in that the
 CC biological activity for reducing L-sorbose is substantially
 CC nullified by gene recombination. The present sequence represents
 CC a Gluconobacter suboxydans L-sorbose reductase peptide, used in
 CC an example from the present invention for generating degenerate
 CC PCR primers.
 CC N.B. This patent is equivalent to the basic NO9901197 in week 199949.
 XX
 SQ Sequence 8 AA;
 Query Match 1.6%; Score 8; DB 20; Length 8;
 Best Local Similarity 100.0%; Pred. No. 7.8e+05;
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 128 MTITEGGY 135
 DB 1 MTITEGGY 8
 RESULT 7
 AAY49915
 ID AAY49915 standard; peptide: 8 AA.
 XX
 AC AAY49915;
 XX
 DT 27-JAN-2000 (first entry)
 XX
 DE Gluconobacter suboxydans L-sorbose reductase peptide SEQ ID NO:4.
 XX
 KW Gluconobacter suboxydans; L-sorbose reductase; genetic engineering;
 KW mutant..
 XX
 OS Gluconobacter suboxydans..
 XX
 PN AU9920390-A.
 PD 23-SEP-1999.
 XX
 PF 11-MAR-1999; 99AU-0020390.
 XX
 PR 13-MAR-1998; 98EP-0104546.
 XX
 PA (HOFF) HOFFMANN LA ROCHE & CO AG F.
 XX
 PI Hoshino T, Tazoe M, Shinjoh M, Kon T;
 XX
 DR WPI; 1999-579276/49.

XX Set of genetically modified mutants not containing L sorbose reductase
 XX Example 3; Fig 2; 33pp; English.
 XX The present invention describes a genetically engineered
 CC microorganism derived from a microorganism belonging to the genus
 CC Gluconobacter or Acetobacter which is characterised in that the
 CC biological activity for reducing L-sorbose is substantially
 CC nullified by gene recombination. The present sequence represents
 CC a Gluconobacter suboxydans L-sorbose reductase peptide, used in
 CC an example from the present invention for generating degenerate
 CC PCR primers.
 CC N.B. This patent is equivalent to the basic NO9901197 in week 199949.
 XX Sequence 8 AA:
 SQ Query Match 1.6%; Score 8; DB 20; Length 8;
 Best Local Similarity 100.0%; Pred. No. 7.8e+05;
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 223 FPNGMVD 230
 Db 1 FPNGMVD 8
 |||||
 |||||
 RESULT 8
 ABP33539
 ID ABP33539 standard; Protein; 103 AA.
 XX AC ABP33539;
 XX 09-JUL-2002 (first entry)
 XX Human reductase-like ORF2512 protein, SEQ ID NO:5024.
 DE Human; ORF; open reading frame; ORFX; drug screening; diagnosis;
 KW disease monitoring; cytokine; cell proliferation; cell differentiation;
 KW immune modulation; haematopoiesis regulation; tissue growth;
 KW angiogenesis; activin; inhibin; chemotactic; chemokinetic; haemostatic;
 KW thrombolytic; tumour inhibition; bodily characteristic; fertility;
 KW behaviour; cancer; proliferative disorder; neurological disorder;
 KW cardiovascular disease; immune system disorder; organ transplantation;
 KW tissue growth disorder; tissue regeneration disorder; diabetes mellitus;
 KW hypothyroidism; cholesterol ester storage disease; infection; vulnery;
 KW vasotropic; antipsoriatic; antidiabetic; cytostatic; neutropic;
 KW neuroprotective; antiatherosclerotic; anticoagulant; thrombolytic;
 KW cardiant; hypotensive; antithyroid; antiinflammatory; immunomodulator;
 KW dermatological; analgesic; virucide; antibacterial; fungicide.
 XX Homo sapiens.
 OS WO200190366-A2.
 PN 29-NOV-2001.
 XX 24-MAY-2001; 2001WO-US17076.
 PF 24-MAY-2000; 2000US-206690P.
 PR (CURA-) CURAGEN CORP.
 XX Leach MD, Shimkets RA;
 PI WPI: 2002-106200/14.
 XX N-PSDB; ABN75565.
 DR Novel human polypeptides and polynucleotides useful for diagnosing,
 PT preventing and treating cardiovascular disease, neurodegenerative,
 PT hyperproliferative disorders and disorders related to organ
 PT transplantation
 XX Claim 10; Page 1509; 2508pp; English.

XX Sequences ABP31028-ABP35561 represent 4534 novel human proteins
 CC designated ORF (open reading frame) 1-4534, and sequences ABN75054-
 CC ABN79587 represent cDNAs encoding them. The invention also encompasses
 CC polypeptides at least 80% identical to the ORF1-ORF4534 (collectively
 CC referred to as ORFX) proteins, polynucleotides at least 85% identical to
 CC the ORFX nucleic acid sequences, vectors and host cells comprising ORFX
 CC polynucleotides, the recombinant production of ORFX proteins, antibodies
 CC specific for ORFX proteins, methods of detecting ORFX polynucleotides and
 CC polypeptides, methods of screening for modulators of ORFX expression or
 CC activity, and methods of screening individuals for a predisposition to an
 CC ORFX-associated disorder. The ORFX proteins of the invention have a wide
 CC range of biological activities, such as cytokine, cell proliferation,
 CC cell differentiation, immune modulation, haematopoiesis regulation,
 CC tissue growth, angiogenesis, activin or inhibin activity, chemotactic/
 CC chemokinetic activity, haemostatic activity, thrombolytic activity,
 CC receptor/ligand, antiinflammatory activity, tumour inhibition activity,
 CC and antiinfective activity, and may also be involved in the determination
 CC of bodily characteristics, fertility and behaviour. ORFX proteins,
 CC nucleic acids and antibodies may be used in the treatment of cancers,
 CC other proliferative disorders such as psoriasis and benign tumours,
 CC neurological disorders such as epilepsy and Alzheimer's disease,
 CC cardiovascular diseases, immune system disorders, disorders related to
 CC organ transplantation, disorders of tissue growth and regeneration,
 CC diseases such as diabetes mellitus, hypothyroidism, and cholesterol ester
 CC storage disease, and infectious diseases caused by viral, bacterial,
 CC fungal and other pathogens. ORFX nucleic acids may also be used as a
 CC source of primers and probes, in the detection of ORFX genomic sequences
 CC or transcripts, in the identification and cloning of homologous
 CC sequences, in genetic diagnosis, and in forensic biology. The ORFX
 CC nucleic acids may additionally be used to produce transgenic animals
 CC which may be useful for studying the function and/or activity of ORFX
 CC protein, and in drug screening. The ORFX proteins may also be used as
 CC immunogens to generate specific antibodies, which are useful in the
 CC diagnosis, treatment and monitoring of ORFX-associated diseases.
 XX Sequence 103 AA:
 SQ Query Match 1.6%; Score 8; DB 23; Length 103;
 Best Local Similarity 100.0%; Pred. No. 11;
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 125 IVSWTITE 132
 Db 79 IVSWTITE 86
 |||||
 |||||
 RESULT 9
 AAY99387
 ID AAY99387 standard; Protein; 163 AA.
 XX AC AAY99387;
 XX 08-AUG-2000 (first entry)
 DT Human PRO1273 (UNQ643) amino acid sequence SEQ ID NO:158.
 XX Human; PRO polypeptide; membrane bound protein; receptor; diagnosis;
 KW transmembrane; secretion; immunoadhesion; pharmaceutical; screening.
 XX Homo sapiens.
 OS WO200012708-A2.
 PN 09-MAR-2000.
 PD 01-SEP-1999; 99WO-US20111.
 XX 01-SEP-1998; 98US-0098716.
 PR 01-SEP-1998; 98US-0098749.
 PR 01-SEP-1998; 98US-0098750.
 PR 02-SEP-1998; 98US-0098803.
 PR 02-SEP-1998; 98US-0098821.

PR	02-SEP-1998;	98US-0098843.	PR	20-OCT-1998;	98US-0105002.
PR	09-SEP-1998;	98US-0099536.	PR	21-OCT-1998;	98US-0105104.
PR	09-SEP-1998;	98US-0099596.	PR	22-OCT-1998;	98US-0105169.
PR	09-SEP-1998;	98US-0099598.	PR	22-OCT-1998;	98US-0105266.
PR	09-SEP-1998;	98US-0099602.	PR	26-OCT-1998;	98US-0105693.
PR	09-SEP-1998;	98US-0099647.	PR	26-OCT-1998;	98US-0105694.
PR	10-SEP-1998;	98US-0099741.	PR	27-OCT-1998;	98US-0105807.
PR	10-SEP-1998;	98US-0099754.	PR	27-OCT-1998;	98US-0105881.
PR	10-SEP-1998;	98US-0099763.	PR	27-OCT-1998;	98US-0105882.
PR	10-SEP-1998;	98US-0099792.	PR	27-OCT-1998;	98US-0106062.
PR	10-SEP-1998;	98US-0099808.	PR	28-OCT-1998;	98US-0106023.
PR	10-SEP-1998;	98US-0099812.	PR	28-OCT-1998;	98US-0106029.
PR	10-SEP-1998;	98US-0099815.	PR	28-OCT-1998;	98US-0106030.
PR	10-SEP-1998;	98US-0099816.	PR	28-OCT-1998;	98US-0106032.
PR	15-SEP-1998;	98US-0100385.	PR	28-OCT-1998;	98US-0106033.
PR	15-SEP-1998;	98US-0100388.	PR	28-OCT-1998;	98US-0106178.
PR	15-SEP-1998;	98US-0100390.	PR	29-OCT-1998;	98US-0106248.
PR	16-SEP-1998;	98US-0100584.	PR	29-OCT-1998;	98US-0106384.
PR	16-SEP-1998;	98US-0100627.	PR	30-OCT-1998;	98US-0108500.
PR	16-SEP-1998;	98US-0100661.	PR	30-OCT-1998;	98US-0106464.
PR	16-SEP-1998;	98US-0100662.	PR	03-NOV-1998;	98US-0106856.
PR	16-SEP-1998;	98US-0100664.	PR	03-NOV-1998;	98US-0106902.
PR	17-SEP-1998;	98US-0100683.	PR	03-NOV-1998;	98US-0106905.
PR	17-SEP-1998;	98US-0100684.	PR	03-NOV-1998;	98US-0106919.
PR	17-SEP-1998;	98US-0100710.	PR	03-NOV-1998;	98US-0106932.
PR	17-SEP-1998;	98US-0100711.	PR	03-NOV-1998;	98US-0106934.
PR	17-SEP-1998;	98US-0100919.	PR	10-NOV-1998;	98US-0107783.
PR	17-SEP-1998;	98US-0100930.	PR	17-NOV-1998;	98US-0108775.
PR	18-SEP-1998;	98US-0100848.	PR	17-NOV-1998;	98US-0108779.
PR	18-SEP-1998;	98US-0100849.	PR	17-NOV-1998;	98US-0108787.
PR	18-SEP-1998;	98US-0101014.	PR	17-NOV-1998;	98US-0108788.
PR	18-SEP-1998;	98US-0101068.	PR	17-NOV-1998;	98US-0108801.
PR	18-SEP-1998;	98US-0101071.	PR	17-NOV-1998;	98US-0108802.
PR	22-SEP-1998;	98US-0101279.	PR	17-NOV-1998;	98US-0108806.
PR	23-SEP-1998;	98US-0101471.	PR	17-NOV-1998;	98US-0108807.
PR	23-SEP-1998;	98US-0101472.	PR	17-NOV-1998;	98US-0108867.
PR	23-SEP-1998;	98US-0101474.	PR	17-NOV-1998;	98US-0108925.
PR	23-SEP-1998;	98US-0101475.	PR	18-NOV-1998;	98US-0108848.
PR	23-SEP-1998;	98US-0101476.	PR	18-NOV-1998;	98US-0108849.
PR	23-SEP-1998;	98US-0101477.	PR	18-NOV-1998;	98US-0108850.
PR	23-SEP-1998;	98US-0101478.	PR	18-NOV-1998;	98US-0108851.
PR	24-SEP-1998;	98US-0101738.	PR	18-NOV-1998;	98US-0108852.
PR	24-SEP-1998;	98US-0101741.	PR	18-NOV-1998;	98US-0108853.
PR	24-SEP-1998;	98US-0101743.	PR	18-NOV-1998;	98US-0108858.
PR	24-SEP-1998;	98US-0101915.	XX		
PR	24-SEP-1998;	98US-0101916.	XX		
PR	29-SEP-1998;	98US-0102207.	XX		
PR	29-SEP-1998;	98US-0102240.	XX		
PR	29-SEP-1998;	98US-0102307.	PA	(GETH) GENENTECH INC.	
PR	29-SEP-1998;	98US-0102330.	PI	Baker K, Goddard A, Gurney AL, Smith V, Watanabe CK, Wood WI;	
PR	29-SEP-1998;	98US-0102331.	XX		
PR	29-SEP-1998;	98US-0102331.	XX		
PR	30-SEP-1998;	98US-0102484.	DR	WPI; 2000-237871/20.	
PR	30-SEP-1998;	98US-0102487.	DR	N-PSDB; AAA37069.	
PR	30-SEP-1998;	98US-0102570.	XX		
PR	30-SEP-1998;	98US-0102571.	PT	New mammalian DNA sequences encoding transmembrane, receptor or	
PR	01-OCT-1998;	98US-0102684.	PT	secreted PRO polypeptides, useful for screening of potential peptide or	
PR	01-OCT-1998;	98US-0102687.	XX	small molecule inhibitors of the relevant receptor/ligand interactions	
PR	02-OCT-1998;	98US-0102965.	PS	Claim 12; Fig 96; 773pp; English.	
PR	06-OCT-1998;	98US-0103258.	XX		
PR	06-OCT-1998;	98US-0103449.	CC	AAA37022 to AAA37144 encode the new isolated human transmembrane,	
PR	07-OCT-1998;	98US-0103314.	CC	receptor or secreted PRO polypeptides given in AA199340 to AA199462. The	
PR	07-OCT-1998;	98US-0103315.	CC	transmembrane and receptor PRO proteins can be used for screening of	
PR	07-OCT-1998;	98US-0103315.	CC	potential peptide or small molecule inhibitors of the relevant	
PR	07-OCT-1998;	98US-0103328.	CC	receptor/ligand interactions. The polypeptides and nucleotide sequences	
PR	07-OCT-1998;	98US-0103395.	CC	encoding then have various industrial applications, including uses as	
PR	07-OCT-1998;	98US-0103396.	CC	pharmaceutical and diagnostic agents. AAA37145 to AAA37330 represent	
PR	07-OCT-1998;	98US-0103401.	CC	PCR primers and hybridisation probes used in the isolation of the PRO	
PR	08-OCT-1998;	98US-0103633.	CC	polypeptides from the present invention.	
PR	08-OCT-1998;	98US-0103678.	XX		
PR	08-OCT-1998;	98US-0103679.	SQ	Sequence 163 AA;	
PR	08-OCT-1998;	98US-0103711.			
PR	14-OCT-1998;	98US-0104257.			
PR	20-OCT-1998;	98US-0104987.			
PR	20-OCT-1998;	98US-0105000.			

Query Match 1.6%; Score 8; DB 21; Length 163;
 Best Local Similarity 100.0%; Pred. No. 17;
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 81 LYSLTETA 88
 Db 136 LYSLTETA 143

RESULT 10
 AAU12423
 ID AAU12423 standard; Protein: 163 AA.
 XX
 AC AAU12423;
 XX
 DT 24-OCT-2001 (first entry)
 XX
 DE Human PRO1273 polypeptide sequence.
 XX
 KW Human secretory and transmembrane; PRO; mammalian; cancer; lung;
 KW breast; prostate; cervical; tumour necrosis factor-alpha; TNF-alpha;
 KW cartilage; ear; proliferation; glucose; free fatty acid; skeletal muscle;
 KW adipocyte; A-peptide; factor VIIA; gene therapy.
 XX
 OS Homo sapiens.
 XX
 PN WO200140466-A2.
 XX
 PD 07-JUN-2001.
 XX
 PF 01-DEC-2000; 2000WO-US32678.
 XX
 PR 01-DEC-1999; 99WO-US28301.
 PR 01-DEC-1999; 99WO-US28634.
 PR 02-DEC-1999; 99WO-US28551.
 PR 02-DEC-1999; 99WO-US28564.
 PR 02-DEC-1999; 99WO-US28565.
 PR 09-DEC-1999; 99US-0170262.
 PR 16-DEC-1999; 99WO-US30095.
 PR 20-DEC-1999; 99WO-US30911.
 PR 20-DEC-1999; 99WO-US30999.
 PR 30-DEC-1999; 99WO-US31243.
 PR 06-JAN-2000; 2000WO-US00277.
 PR 11-FEB-2000; 2000WO-US00376.
 PR 11-FEB-2000; 2000WO-US03565.
 PR 18-FEB-2000; 2000WO-US04341.
 PR 18-FEB-2000; 2000WO-US04342.
 PR 22-FEB-2000; 2000WO-US04414.
 PR 24-FEB-2000; 2000WO-US04914.
 PR 24-FEB-2000; 2000WO-US05004.
 PR 01-MAR-2000; 2000WO-US05601.
 PR 20-MAR-2000; 2000WO-US07377.
 PR 21-MAR-2000; 2000WO-US07532.
 PR 30-MAR-2000; 2000WO-US08439.
 PR 17-MAY-2000; 2000WO-US13705.
 PR 22-MAY-2000; 2000WO-US14042.
 PR 30-MAY-2000; 2000WO-US14941.
 PR 02-JUN-2000; 2000WO-US15264.
 PR 10-NOV-2000; 2000WO-US30873.
 XX
 PA (GETH) GENENTECH INC.
 XX
 PI Baker KP, Beresini M, Deforge L, Desnoyers L, Filvaroff E, Gao W;
 PI Gerritsen ME, Goddard A, Godowski PJ, Gurney AL, Sherwood S;
 PI Smith V, Stewart TA, Tumas D, Watanabe CK, Wood WI, Zhang Z;
 XX
 WPI: 2001-408281/43.
 DR N-PSDB; AAS21495.
 XX
 XX Isolated, secretory and transmembrane PRO polypeptide used to detect
 PT other PRO polypeptides, link bioactive molecules to cells expressing
 PT PRO polypeptides, and detect the presence of mammalian tumours e.g.
 PT lung, breast, prostate, cervical
 XX
 PS Claim 12; Fig 504; 813pp; English.
 XX

CC AAU12172-AAU12446 represent novel human secretory and transmembrane
 CC PRO polypeptides. The PRO polypeptides are useful to detect other
 CC PRO polypeptides, to link bioactive molecules to cells expressing
 CC PRO polypeptides, to modulate biological activities of cells expressing
 CC PRO polypeptides, and to detect the presence of mammalian lung, colon,
 CC breast, prostate, rectal, cervical or liver tumours by comparing PRO
 CC polypeptide expression in a cell sample to that in a control sample.
 CC Some of the 275 sequences are also useful to stimulate the release of
 CC tumour necrosis factor-alpha (TNF-alpha) from human blood, the
 CC proliferation or differentiation of chondrocytes, the proliferation or
 CC gene expression in pericyte cells, the release of proteoglycans from
 CC cartilage, the proliferation of inner ear utricular supporting cells or
 CC of T-lymphocytes, the release of a cytokine from peripheral blood
 CC monocytes (PBMCs), or the proliferation of endothelial cells. Some of
 CC the PRO polypeptides may modulate glucose or free fatty acid uptake by
 CC skeletal muscle cells or by adipocytes; or inhibit binding of A-peptide
 CC to factor VIIA. The PRO polypeptides can be used in assays to identify
 CC molecules involved in binding interactions. The polynucleotides encoding
 CC PRO polypeptides can be used to generate probes, antisense RNA/DNA,
 CC transgenic or knock out animals and can be used in gene therapy.
 XX
 SQ Sequence 163 AA;

Query Match 1.6%; Score 8; DB 22; Length 163;
 Best Local Similarity 100.0%; Pred. No. 17;
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 81 LYSLTETA 88
 Db 136 LYSLTETA 143

RESULT 11
 AAB66136
 ID AAB66136 standard; protein; 163 AA.
 XX
 AC AAB66136;
 XX
 DT 02-APR-2001 (first entry)
 XX
 DE Protein of the invention #48.
 XX
 KW Secreted; transmembrane; gene therapy.
 XX
 OS Unidentified.
 XX
 PN WO200078961-A1.
 XX
 PD 28-DEC-2000.
 XX
 PF 18-FEB-2000; 2000WO-US04342.
 XX
 PR 23-JUN-1999; 99US-0141037.
 PR 20-JUL-1999; 99US-0144758.
 PR 26-JUL-1999; 99US-0145698.
 PR 01-SEP-1999; 99WO-US20111.
 PR 29-OCT-1999; 99US-0162506.
 PR 30-NOV-1999; 99WO-US28313.
 PR 02-DEC-1999; 99WO-US28551.
 PR 16-DEC-1999; 99WO-US30095.
 PR 05-JAN-2000; 2000WO-US00219.
 PR 06-JAN-2000; 2000WO-US00376.
 XX
 PA (GETH) GENENTECH INC.
 XX
 PI Baker KP, Botstein D, Desnoyers L, Eaton DL, Ferrara N, Fong S;
 PI Gao W, Goddard A, Godowski PJ, Grimaldi CJ, Gurney AL, Hillan KJ;
 PI Pan J, Paoni NF, Roy MA, Smith V, Stewart TA, Tumas D;
 PI Watanabe CK, Williams PM, Wood WI;
 XX
 WPI: 2001-071395/08.
 DR
 XX Secreted and transmembrane proteins and nucleic acids designated PRO,

PT useful as hybridization probes, in chromosome and gene mapping and gene
PT therapy -

XX Claim 1; Fig 96; 787pp; English.

XX The present invention relates to secreted and transmembrane proteins.
CC These proteins and the DNA encoding them may be used as hybridization
CC probes, in chromosome and gene mapping and in the generation of
CC anti-sense RNA and DNA. They may also be used to generate either
CC transgenic animals or knockout animals which are in turn useful for
CC development and screening of therapeutically useful reagents.
CC The nucleic acids may also be used in gene therapy.

XX Sequence 163 AA;

Query Match 1.6%; Score 8; DB 22; Length 163;
Best Local Similarity 100.0%; Pred. No. 17;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy 81 LYSLTETA 88
| | | | | | | |
Db 136 LYSLTETA 143

RESULT 12
AAB36671
ID AAB36671 standard; Protein; 163 AA.

XX AAB36671;

XX 14-MAR-2001 (first entry)

XX Human secretory protein TGC-715 SEQ ID NO:11.

DE Human; secretory protein; cancer; immune disease; infectious disease;
KW lung function disorder; liver function disorder; antiinflammatory;
KW gastrointestinal disorder; cytostatic; haematopoietic; anticoagulant;
KW immunomodulatory; hepatotropic; cell proliferation-stimulant;
KW cell migratory agent; cell differentiation-inducer.

XX Homo sapiens.

XX WO200071581-A1.

XX 30-NOV-2000.

XX 19-MAY-2000; 2000WO-JP03221.

XX 20-MAY-1999; 99JP-0140229.

XX (TAKE) TAKEDA CHEM IND LTD.

XX Itoh Y, Mogi S, Tanaka H, Ohkubo S, Ogi K;

XX WPI; 2001-032023/04.

XX N-PSDB; AAC90711.

XX Novel secretory protein and its salt with e.g. anti-cancer,
PT anti-inflammatory and hematopoietic, effects, applicable as drugs in
PT remedies and preventives to treat diseases like cancer and immune
PT diseases -

XX Claim 1; Page 92-93; 122pp; Japanese.

XX AAC90701 to AAC90715 encode the human secretory proteins given in
CC AAB36661 to AAB36675. The proteins can have cytostatic,
CC anti-inflammatory, haematopoietic, anti-coagulant, immunomodulatory and
CC hepatotropic activities, and can be used as cell migratory agents, cell
CC proliferation-stimulants and cell differentiation-inducers. The proteins
CC are useful in the treatment and prevention of diseases such as cancer,
CC lung function disorder, liver function disorder, gastrointestinal
CC disorder and immune diseases. AAC90716 to AAC90755 represent PCR primers
CC which are used in the exemplification of the present invention.

XX Sequence 163 AA;

Query Match 1.6%; Score 8; DB 22; Length 163;
Best Local Similarity 100.0%; Pred. No. 17;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy 81 LYSLTETA 88
| | | | | | | |
Db 136 LYSLTETA 143

RESULT 13
ABB95524
ID ABB95524 standard; Protein; 163 AA.

XX ABB95524;

XX 19-JUL-2002 (first entry)

XX Human angiogenesis related protein PRO1273 SEQ ID NO: 204.

XX Human; angiogenesis; PRO protein; cardiovascularisation; wound; cancer;
KW atherosclerosis; cardiac hypertrophy; gene therapy; endothelial disorder;
KW cardiant; cytostatic; antiangiogenic; hypotensive; vulnery;

XX antiarteriosclerotic.

XX Homo sapiens.

XX WO200208284-A2.

XX 31-JAN-2002.

XX 09-JUL-2001; 2001WO-US21735.

XX 20-JUL-2000; 2000US-219556P.

XX 25-JUL-2000; 2000US-220624P.

XX 25-JUL-2000; 2000US-220664P.

XX 28-JUL-2000; 2000WO-US20710.

XX 02-AUG-2000; 2000US-222695P.

XX 17-AUG-2000; 2000US-0643657.

XX 23-AUG-2000; 2000WO-US23522.

XX 24-AUG-2000; 2000WO-US23328.

XX 07-SEP-2000; 2000US-230978P.

XX 15-SEP-2000; 2000US-000000P.

XX 18-SEP-2000; 2000US-0664610.

XX 18-SEP-2000; 2000US-0665350.

XX 24-OCT-2000; 2000US-242922P.

XX 08-NOV-2000; 2000US-0709238.

XX 10-NOV-2000; 2000WO-US30952.

XX 01-DEC-2000; 2000WO-US32678.

XX 20-DEC-2000; 2000US-0747259.

XX 20-DEC-2000; 2000WO-US34956.

XX 22-JAN-2001; 2001US-0767609.

XX 28-FEB-2001; 2001US-0796498.

XX 01-MAR-2001; 2001WO-US06520.

XX 09-MAR-2001; 2001US-0802706.

XX 14-MAR-2001; 2001US-0808689.

XX 22-MAR-2001; 2001US-0816744.

XX 05-APR-2001; 2001US-0828366.

XX 10-MAY-2001; 2001US-0854208.

XX 10-MAY-2001; 2001US-0854280.

XX 25-MAY-2001; 2001US-0866028.

XX 25-MAY-2001; 2001US-0866034.

XX 30-MAY-2001; 2001WO-US17092.

XX 30-MAY-2001; 2001US-0870574.

XX 01-JUN-2001; 2001WO-US17443.

XX 20-JUN-2001; 2001WO-US17800.

XX 28-JUN-2001; 2001WO-US19692.

XX 28-JUN-2001; 2001WO-US00000.

```

PA (GETH ) GENENTECH INC.
PA (BAKE/) BAKER K P.
PA (FERR/) FERRARA N.
PA (GERB/) GERBER H.
PA (GERR/) GERRITSEN M E.
PA (GODD/) GODDARD A.
PA (GODO/) GODOWSKI P J.
PA (GURN/) GURNEY A L.
PA (HILL/) HILLAN K J.
PA (MARS/) MARSTERS S A.
PA (PANJ/) PAN J.
PA (PAONI/) PAONI N F.
PA (STEP/) STEPHAN J F.
PA (WATA/) WATANABE C K.
PA (WILL/) WILLIAMS P M.
PA (WOOD/) WOOD W I.
XX
PI Baker KP, Ferrara N, Gerber H, Gerritsen ME, Goddard A;
PI Godowski PJ, Gurney AL, Hillan KJ, Marsters SA, Pan J, Paoni NF;
PI Stephan JF, Watanabe CK, Williams PM, Wood WI, Ye W;
XX
DR WPI: 2002-171999/22.
DR N-PSDB; ABL95662.
XX
PT One hundred and eighty seven nucleic acids encoding PRO polypeptides,
PT useful in diagnosis and treatment of cardiovascular (e.g. myocardial
PT infarction), endothelial or angiogenic disorders in a mammal -
XX
PS Claim 11; Fig 204; 567pp; English.
XX
CC The present invention provides the protein and coding sequences of human
CC PRO proteins. These are useful for treating or diagnosing a
CC cardiovascular, endothelial or angiogenic disorder, including cardiac
CC hypertrophy, trauma, cancer, age-related macular degeneration,
CC atherosclerosis, hypertension, arterial restenosis, rheumatoid arthritis,
CC angina, myocardial infarctions, thrombophlebitis, lymphangiitis, tumour
CC angiogenesis (such as breast carcinoma and liver carcinoma) and wound
CC healing. The present sequence is a PRO protein of the invention.
XX
SQ Sequence 163 AA;

Query Match 1.6%; Score 8; DB 23; Length 163;
Best Local Similarity 100.0%; Pred. No. 17;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 81 LYSLETETA 88
DB 136 LYSLETETA 143
|||||||

RESULT 14
ABB84918
ID ABB84918 standard; Protein; 163 AA.
XX
AC ABB84918;
XX
DT 16-MAY-2002 (first entry)
XX
DE Human PRO1273 protein sequence SEQ ID NO:204.
XX
KW Human; angiogenesis; cardiac; cytostatic; antiangiogenic; hypotensive;
KW vulnery; antiarteriosclerotic; PRO agonist; PRO antagonist; trauma;
KW gene therapy; cardiovascular disorder; endothelial disorder; cancer;
KW angiogenic disorder; cardiac hypertrophy; atherosclerosis; hypertension;
KW age-related macular degeneration; arterial restenosis; angina;
KW rheumatoid arthritis; myocardial infarction; thrombophlebitis;
KW lymphangiitis; tumour angiogenesis; breast carcinoma; liver carcinoma;
KW wound healing; chromosome mapping; gene mapping.
XX
OS Homo sapiens.
XX
PN WO200200690-A2.
XX

```

```

PD 03-JAN-2002.
XX
PF 20-JUN-2001; 2001WO-US19692.
XX
PR 23-JUN-2000; 2000US-213637P.
PR 20-JUL-2000; 2000US-219556P.
PR 25-JUL-2000; 2000US-220624P.
PR 25-JUL-2000; 2000US-220664P.
PR 28-JUL-2000; 2000WO-US20710.
PR 02-AUG-2000; 2000US-222695P.
PR 17-AUG-2000; 2000US-0643657.
PR 23-AUG-2000; 2000WO-US23522.
PR 24-AUG-2000; 2000WO-US23328.
PR 07-SEP-2000; 2000US-230978P.
PR 18-SEP-2000; 2000US-0664610.
PR 18-SEP-2000; 2000US-0665350.
PR 24-OCT-2000; 2000US-242922P.
PR 08-NOV-2000; 2000US-070923B.
PR 08-NOV-2000; 2000WO-US30952.
PR 10-NOV-2000; 2000WO-US30873.
PR 01-DEC-2000; 2000WO-US32678.
PR 20-DEC-2000; 2000US-0747259.
PR 20-DEC-2000; 2000WO-US34956.
PR 22-JAN-2001; 2001US-0767609.
PR 28-FEB-2001; 2001US-0796498.
PR 28-FEB-2001; 2001WO-US06520.
PR 01-MAR-2001; 2001WO-US06666.
PR 09-MAR-2001; 2001US-0802706.
PR 14-MAR-2001; 2001US-0808689.
PR 22-MAR-2001; 2001US-0816744.
PR 05-APR-2001; 2001US-0828366.
PR 10-MAY-2001; 2001US-0854208.
PR 10-MAY-2001; 2001US-0854280.
PR 25-MAY-2001; 2001US-0866028.
PR 25-MAY-2001; 2001US-0866034.
PR 25-MAY-2001; 2001WO-US17092.
PR 30-MAY-2001; 2001US-0870574.
PR 30-MAY-2001; 2001WO-US17443.
PR 01-JUN-2001; 2001WO-US17800.
XX
XX (GETH ) GENENTECH INC.
XX
PI Baker KP, Ferrara N, Gerber H, Gerritsen ME, Goddard A;
PI Godowski PJ, Gurney AL, Hillan KJ, Marsters SA, Pan J, Paoni NF;
PI Stephan JF, Watanabe CK, Williams PM, Wood WI, Ye W;
XX
XX WPI: 2002-090516/12.
XX N-PSDB; ABL88173.
XX
PT One hundred and eighty seven nucleic acids encoding PRO polypeptides,
PT useful in diagnosis and treatment of cardiovascular (e.g. myocardial
PT infarction), endothelial or angiogenic disorders in a mammal -
XX
PS Claim 11; Fig 204; 565pp; English.
XX
CC ABL88072 to ABL88258 encode the PRO proteins given in ABB84817 to
CC ABB85003. The PRO proteins and polynucleotides have cardiac, cytostatic,
CC antiangiogenic, hypotensive, vulnery and antiarteriosclerotic
CC activities, and can be used in gene therapy. The PRO polynucleotides,
CC proteins, agonists and antagonists are useful for treating or diagnosing
CC a cardiovascular, endothelial or angiogenic disorder in a mammal,
CC e.g. cardiac hypertrophy, trauma, cancer, age-related macular
CC degeneration, atherosclerosis, hypertension, arterial restenosis,
CC rheumatoid arthritis, angina, myocardial infarctions, thrombophlebitis,
CC lymphangiitis, tumour angiogenesis (such as breast carcinoma and liver
CC carcinoma) and wound healing. The PRO polynucleotides have applications
CC in molecular biology, including use as hybridisation probes, and in
CC chromosome and gene mapping. ABL88259 to ABL88267 represent primers and
CC probes used in the exemplification of the present invention.
XX
SQ Sequence 163 AA;

Query Match 1.6%; Score 8; DB 23; Length 163;

```

Best Local Similarity 100.0%; Pred. No. 17;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 81 LYSLTETA 88
DB 136 LYSLTETA 143

RESULT 15
AAU83683
ID AAU83683 standard; Protein; 163 AA.
XX
AC AAU83683;
XX
DT 08-MAY-2002 (first entry)
XX
DE Human PRO protein, Seq ID No 184.
XX
KW Human; secreted protein; PRO; tumour; lung cancer; colon cancer;
KW breast cancer; prostate tumour; rectal tumour; liver tumour;
KW pericyte cell proliferation; chondrocyte cell proliferation;
KW tumour necrosis factor-alpha.
XX
OS Homo sapiens.
XX
PN WO200208288-A2.
XX
PD 31-JAN-2002.
XX
PF 29-JUN-2001; 2001WO-US21066.
XX
PR 20-JUL-2000; 2000US-219556P.
PR 25-JUL-2000; 2000US-220585P.
PR 25-JUL-2000; 2000US-220605P.
PR 25-JUL-2000; 2000US-220607P.
PR 25-JUL-2000; 2000US-220624P.
PR 25-JUL-2000; 2000US-220638P.
PR 25-JUL-2000; 2000US-220664P.
PR 25-JUL-2000; 2000US-220666P.
PR 26-JUL-2000; 2000US-220893P.
PR 28-JUL-2000; 2000WO-US20710.
PR 23-AUG-2000; 2000WO-US23522.
PR 24-AUG-2000; 2000WO-US23328.
PR 15-SEP-2000; 2000US-000000P.
PR 10-NOV-2000; 2000WO-US30873.
PR 28-NOV-2000; 2000US-253646P.
PR 01-DEC-2000; 2000WO-US32678.
PR 20-DEC-2000; 2000US-0747259.
PR 20-DEC-2000; 2000WO-US34956.
PR 28-FEB-2001; 2001WO-US06520.
PR 10-MAY-2001; 2001US-0854280.
PR 25-MAY-2001; 2001WO-US17092.
XX
PA (GETH) GENENTECH INC.
XX
PI Baker KP, Desnoyers L, Gerritsen ME, Goddard A, Godowski PJ;
PI Grimaldi JC, Gurney AL, Smith V, Stephan JF, Watanabe CK, Wood WL;
XX
DR WPI; 2002-172001/22.
DR N-P5DB; ABR33627.
XX
PT One hundred and twenty two nucleic acids encoding PRO polypeptides,
PT useful for treating a PRO related disorder and for diagnosing tumours
PT such as lung cancer, colon cancer, breast tumour, prostate tumour, rectal
PT tumour or liver tumour -
XX
PS Claim 11; Figure 184; 359pp; English.
XX
CC The invention relates to one hundred and twenty two nucleic acids
CC encoding PRO polypeptides. The sequences of the 122 PRO polynucleotides
CC encode human secreted proteins. The PRO nucleic acids, polypeptides,
CC agonists and antagonists are useful for treating a PRO related disorder.
CC The PRO polypeptides are useful for diagnosing tumours, especially lung

cancer, colon cancer, breast tumour, prostate tumour, rectal tumour or
liver tumour. The PRO polypeptides are useful for stimulating the
proliferation of, or gene expression, in pericyte cells, for stimulating
the proliferation or differentiation of chondrocyte cells, for
stimulating the release of tumour necrosis factor-alpha from human blood,
for stimulating or inhibiting the proliferation of normal human dermal
fibroblast cells. The PRO polypeptide may also be used as molecular
weight markers and for tissue typing. The PRO nucleic acids have
applications in molecular biology, including use as hybridisation probes,
and in chromosome and gene mapping. AAU83592-AAU83713 represent human PRO
protein sequences of the invention.

XX
SQ Sequence 163 AA;
XX

Query Match 1.6%; Score 8; DB 23; Length 163;
Best Local Similarity 100.0%; Pred. No. 17;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 81 LYSLTETA 88
DB 136 LYSLTETA 143

Search completed: May 8, 2003, 16:42:50
Job time : 57 secs

GenCore version 5.1.4_p5_4578
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: May 8, 2003, 16:31:32 ; Search time 17 Seconds
(without alignments)
839.419 Million cell updates/sec

Title: US-09-926-163B-2

Perfect score: 485

Sequence: 1 MITRETLKSLPANVQAPPDY.....KVIVLRKIIRKGVKAAIPA 485

Scoring table: OLIGO

Gapop 60.0 , Gapext 60.0

Searched: 262574 seqs, 29422922 residues

Word size : 0

Total number of hits satisfying chosen parameters: 262574

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Database :

Issued_Patents_AA.*

1: /cgn2_6/ptodata/1/iaa/5A_COMB.pap.*

2: /cgn2_6/ptodata/1/iaa/5B_COMB.pap.*

3: /cgn2_6/ptodata/1/iaa/6A_COMB.pap.*

4: /cgn2_6/ptodata/1/iaa/6B_COMB.pap.*

5: /cgn2_6/ptodata/1/iaa/PCTUS_COMB.pap.*

6: /cgn2_6/ptodata/1/iaa/backfiles1.pap.*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	7	1.4	25	1	US-07-966-187-4
2	7	1.4	25	1	US-08-371-121-3
3	7	1.4	25	3	US-07-927-381-5
4	7	1.4	126	1	US-08-202-389-10
5	7	1.4	266	2	US-08-685-992-12
6	7	1.4	266	2	US-09-144-925-12
7	7	1.4	303	2	US-08-286-819A-23
8	7	1.4	303	3	US-08-980-357-23
9	7	1.4	366	1	US-08-481-377-6
10	7	1.4	366	3	US-08-827-336-9
11	7	1.4	366	3	US-09-153-733A-6
12	7	1.4	366	4	US-09-389-705-6
13	7	1.4	366	4	US-09-357-905-9
14	7	1.4	366	5	PCT-US94-00666-6
15	7	1.4	377	4	US-09-269-137-4
16	7	1.4	401	2	US-08-741-327E-11
17	7	1.4	401	4	US-09-064-935-11
18	7	1.4	404	3	US-08-844-065-2
19	7	1.4	424	2	US-08-920-634-2
20	7	1.4	455	4	US-09-308-003-13
21	7	1.4	492	3	US-08-724-466B-2
22	7	1.4	492	4	US-08-882-164D-2
23	7	1.4	497	3	US-08-724-466B-4
24	7	1.4	497	4	US-08-882-164D-4
25	7	1.4	497	4	US-08-882-164D-32
26	7	1.4	593	1	US-08-202-389-12
27	7	1.4	593	1	US-08-018-129-5

28	7	1.4	593	2	US-08-448-250-5	Sequence 5, Appli
29	7	1.4	968	1	US-08-426-236-2	Sequence 2, Appli
30	7	1.4	1684	3	US-08-665-259-25	Sequence 25, Appli
31	7	1.4	1684	3	US-08-762-500-25	Sequence 25, Appli
32	7	1.4	1704	3	US-08-762-500-75	Sequence 75, Appli
33	7	1.4	2291	2	US-08-286-819A-29	Sequence 29, Appli
34	7	1.4	2291	3	US-08-980-357-29	Sequence 29, Appli
35	7	1.4	5087	4	US-09-144-085-1	Sequence 1, Appli
36	7	1.4	6095	4	US-09-144-085-2	Sequence 2, Appli
37	6	1.2	11	3	US-08-159-339A-1157	Sequence 1157, Ap
38	6	1.2	14	1	US-08-440-391-11	Sequence 11, Appli
39	6	1.2	14	2	US-08-908-507A-11	Sequence 11, Appli
40	6	1.2	14	4	US-09-236-385A-11	Sequence 11, Appli
41	6	1.2	14	5	PCT-US96-06122-11	Sequence 11, Appli
42	6	1.2	20	1	US-08-321-071A-19	Sequence 19, Appli
43	6	1.2	34	1	US-08-014-153D-34	Sequence 34, Appli
44	6	1.2	35	4	US-08-942-686-6	Sequence 6, Appli
45	6	1.2	43	1	US-07-998-003A-73	Sequence 73, Appli

ALIGNMENTS

RESULT 1

US-07-966-187-4
; Sequence 4, Application US/07966187
; Patent No. 5477001

; GENERAL INFORMATION:

; APPLICANT: SASS, Catherine

; APPLICANT: LEGUAY, Jean-Jacques

; APPLICANT: GRISON, Rene

; APPLICANT: TOPPAN, Alain

; TITLE OF INVENTION: Recombinant DNA coding for a novel

; TITLE OF INVENTION: protein having beta 1.3 glucanase activity, bacteria

; TITLE OF INVENTION: containing this DNA, transformed plant cells and plants.

; NUMBER OF SEQUENCES: 26

; CORRESPONDENCE ADDRESS:

; ADDRESSEE: FOLEY & LARDNER

; STREET: King Street Station, Suite 500, 1800 Diagonal

; STREET: Road, PO Box 299

; CITY: ALEXANDRIA

; STATE: VIRGINIA

; COUNTRY: USA

; ZIP: 22313-0299

; COMPUTER READABLE FORM:

; MEDIUM TYPE: Floppy disk

; COMPUTER: IBM PC compatible

; OPERATING SYSTEM: PC-DOS/MS-DOS

; SOFTWARE: Patent In Release #1.0, Version #1.25

; CURRENT APPLICATION DATA:

; APPLICATION NUMBER: US/07/966.187

; FILING DATE: 19930125

; CLASSIFICATION: 800

; PRIOR APPLICATION DATA:

; APPLICATION NUMBER: PCT/FR92/00268

; FILING DATE: 25-MAR-1992

; PRIOR APPLICATION DATA:

; APPLICATION NUMBER: FR 91/03588

; FILING DATE: 25-MAR-1991

; NAME: SAXE, Bernhard D

; REGISTRATION NUMBER: 28,665

; REFERENCE/DOCKET NUMBER: 16781/381

; TELECOMMUNICATION INFORMATION:

; TELEPHONE: (703) 836-9300

; TELEFAX: (703) 683-4109

; TELEX: 899149

; INFORMATION FOR SEQ ID NO: 4:

; SEQUENCE CHARACTERISTICS:

; LENGTH: 25 amino acids

; TYPE: AMINO ACID

; TOPOLOGY: linear

; MOLECULE TYPE: peptide

US-07-966-187-4

Query Match 1.4%; Score 7; DB 1; Length 25;
Best Local Similarity 100.0%; Pred. No. 9.8;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 88 APSGKST 94
|||||
Db 2 APSGKST 8

RESULT 2

US-08-371-121-3
; Sequence 3, Application US/08371121
; Patent No. 5652123
; GENERAL INFORMATION:
; APPLICANT: CAPUT, Daniel
; APPLICANT: FERRARA, Pascal
; APPLICANT: GUILLEMOT, Jean-Claude
; APPLICANT: LEPLATOIS, Pascal
; APPLICANT: MINTY, Adrian
; APPLICANT: KAGHAD, Mourad
; APPLICANT: LABIT-LE BOUTEILLER, Christine
; APPLICANT: MAGAZIN, Marilyn
; TITLE OF INVENTION: Protein having a cytokine type
; TITLE OF INVENTION: activity, recombinant DNA coding for this protein,
; TITLE OF INVENTION: transformed cells and microorganisms.
; NUMBER OF SEQUENCES: 27
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: FOLEY & LARDNER
; STREET: 3000 K Street, N.W., Suite 500
; CITY: Washington, D.C.
; COUNTRY: USA
; ZIP: 20007-5109

COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/371,121
FILING DATE:
CLASSIFICATION: 435

PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/938,161
FILING DATE: 30-NOV-1992
PRIOR APPLICATION DATA:
APPLICATION NUMBER: PCT/FR92/00280
FILING DATE: 27-MAR-1992
PRIOR APPLICATION DATA:
APPLICATION NUMBER: FR 91 00137
FILING DATE: 08-JAN-1992
PRIOR APPLICATION DATA:
APPLICATION NUMBER: FR 91 03904
FILING DATE: 29-MAR-1991
ATTORNEY/AGENT INFORMATION:
NAME: SAXE, Bernhard D.
REGISTRATION NUMBER: 28,665
REFERENCE/DOCKET NUMBER: 16781/383
TELECOMMUNICATION INFORMATION:
TELEPHONE: (202) 672-5300
TELEFAX: (202) 672-5399
TELEX: 904136

INFORMATION FOR SEQ ID NO: 3:
SEQUENCE CHARACTERISTICS:
LENGTH: 25 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: peptide

US-08-371-121-3

Query Match 1.4%; Score 7; DB 1; Length 25;
Best Local Similarity 100.0%; Pred. No. 9.8;

Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 88 APSGKST 94
|||||
Db 2 APSGKST 8

RESULT 3

US-07-927-391-5
; Sequence 5, Application US/07927391
; Patent No. 6001649
; GENERAL INFORMATION:
; APPLICANT: CAPUT, Daniel
; APPLICANT: FERRARA, Pascal
; APPLICANT: MILOUX, Brigitte
; APPLICANT: MINTY, Adrian
; APPLICANT: VITA, Natalio
; TITLE OF INVENTION: Protein having a cytokin type
; TITLE OF INVENTION: activity, and recombinant DNA, expression vector and hosts
; TITLE OF INVENTION: for its preparation.
; NUMBER OF SEQUENCES: 25
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: FOLEY & LARDNER
; STREET: King Street Station, Suite 500, 1800 Diagonal
; STREET: Road, PO Box 299
; CITY: ALEXANDRIA
; STATE: VIRGINIA
; COUNTRY: USA
; ZIP: 22313-0299

COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/07/927,391
FILING DATE: 19920929
CLASSIFICATION: 530

ATTORNEY/AGENT INFORMATION:
NAME: SAXE, Bernhard D.
REGISTRATION NUMBER: 28,665
REFERENCE/DOCKET NUMBER: 16781/369
TELECOMMUNICATION INFORMATION:
TELEPHONE: (703) 836-9300
TELEFAX: (703) 683-4109
TELEX: 899149

INFORMATION FOR SEQ ID NO: 5:
SEQUENCE CHARACTERISTICS:
LENGTH: 25 amino acids
TYPE: AMINO ACID
TOPOLOGY: linear
MOLECULE TYPE: peptide

US-07-927-391-5

Query Match 1.4%; Score 7; DB 3; Length 25;
Best Local Similarity 100.0%; Pred. No. 9.8;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 88 APSGKST 94
|||||
Db 2 APSGKST 8

RESULT 4

US-08-202-389-10
; Sequence 10, Application US/08202389
; Patent No. 5536636
; GENERAL INFORMATION:
; APPLICANT: Freeman Jr., Robert M.
; APPLICANT: Plutsky, Jorge
; APPLICANT: Neel, Benjamin G.
; APPLICANT: Rosenberg, Robert D.
; TITLE OF INVENTION: IDENTIFICATION OF NOVEL TYROSINE

;; TITLE OF INVENTION: PHOSPHATASES HAVING SH2 DOMAINS
;; NUMBER OF SEQUENCES: 54
;; CORRESPONDENCE ADDRESS:
;; ADDRESSEE: Hamilton, Brook, Smith & Reynolds, P.C.
;; STREET: Two Militia Drive
;; CITY: Lexington
;; STATE: MA
;; COUNTRY: USA
;; ZIP: 02173
;; COMPUTER READABLE FORM:
;; MEDIUM TYPE: Floppy disk
;; COMPUTER: IBM PC Compatible
;; OPERATING SYSTEM: PC-DOS/MS-DOS
;; SOFTWARE: PatentIn Release #1.0, Version #1.25
;; CURRENT APPLICATION DATA:
;; APPLICATION NUMBER: US/08/202,389
;; FILING DATE: 28-FEB-1994
;; CLASSIFICATION: 435
;; PRIOR APPLICATION DATA:
;; APPLICATION NUMBER: US 07/983,926
;; FILING DATE: 01-DEC-1992
;; PRIOR APPLICATION DATA:
;; APPLICATION NUMBER: US 07/829,141
;; FILING DATE: 31-JAN-1992
;; PRIOR APPLICATION DATA:
;; APPLICATION NUMBER: US 07/721,112
;; FILING DATE: 26-JUN-1991
;; ATTORNEY/AGENT INFORMATION:
;; NAME: Granahan, Patricia
;; REGISTRATION NUMBER: 32,227
;; REFERENCE/DOCKET NUMBER: BTH92-05WA
;; TELEPHONE: (617) 861-6240
;; TELEFAX: (617) 861-9540
;; INFORMATION FOR SEQ ID NO: 10:
;; SEQUENCE CHARACTERISTICS:
;; LENGTH: 126 amino acids
;; TYPE: amino acid
;; TOPOLOGY: linear
;; MOLECULE TYPE: protein
US-08-202-389-10

Query Match 1.4%; Score 7; DB 1; Length 126;
Best Local Similarity 100.0%; Pred. No. 41;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 473 IIREKGV 479
Db 113 IIREKGV 119

RESULT 5
US-08-685-992-12
;; Sequence 12, Application US/08685992
;; Patent No. 5912138
;; GENERAL INFORMATION:
;; APPLICANT: Tonks, Nicholas
;; APPLICANT: Flint, Andrew J.
;; TITLE OF INVENTION: SUBSTRATE TRAPPING PROTEIN
;; TITLE OF INVENTION: TYROSINE PHOSPHATASES
;; NUMBER OF SEQUENCES: 36
;; CORRESPONDENCE ADDRESS:
;; ADDRESSEE: HAMILTON, BROOK, SMITH & REYNOLDS, P.C.
;; STREET: Two Militia Drive
;; CITY: Lexington
;; STATE: MA
;; COUNTRY: USA
;; ZIP: 02173
;; COMPUTER READABLE FORM:
;; MEDIUM TYPE: Diskette
;; COMPUTER: IBM Compatible
;; OPERATING SYSTEM: Windows 95
;; SOFTWARE: FastSEQ for Windows Version 2.0b

;; CURRENT APPLICATION DATA:
;; APPLICATION NUMBER: US/08/685,992
;; FILING DATE: 25-JUL-1996
;; CLASSIFICATION: 435
;; PRIOR APPLICATION DATA:
;; APPLICATION NUMBER:
;; FILING DATE:
;; ATTORNEY/AGENT INFORMATION:
;; NAME: Granahan, Patricia
;; REGISTRATION NUMBER: 32,227
;; REFERENCE/DOCKET NUMBER: CSHL96-03
;; TELEPHONE: 781-861-6240
;; TELEFAX: 781-861-9540
;; TELEX:
;; INFORMATION FOR SEQ ID NO: 12:
;; SEQUENCE CHARACTERISTICS:
;; LENGTH: 266 amino acids
;; TYPE: amino acid
;; STRANDEDNESS: single
;; TOPOLOGY: linear
;; MOLECULE TYPE: peptide
US-08-685-992-12

Query Match 1.4%; Score 7; DB 2; Length 266;
Best Local Similarity 100.0%; Pred. No. 81;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 473 IIREKGV 479
Db 215 IIREKGV 221

RESULT 6
US-09-144-925-12
;; Sequence 12, Application US/09144925
;; Patent No. 5951979
;; GENERAL INFORMATION:
;; APPLICANT: Tonks, Nicholas
;; APPLICANT: Flint, Andrew J.
;; TITLE OF INVENTION: SUBSTRATE TRAPPING PROTEIN
;; TITLE OF INVENTION: TYROSINE PHOSPHATASES
;; NUMBER OF SEQUENCES: 36
;; CORRESPONDENCE ADDRESS:
;; ADDRESSEE: HAMILTON, BROOK, SMITH & REYNOLDS, P.C.
;; STREET: Two Militia Drive
;; CITY: Lexington
;; STATE: MA
;; COUNTRY: USA
;; ZIP: 02421-4799
;; COMPUTER READABLE FORM:
;; MEDIUM TYPE: Diskette
;; COMPUTER: IBM Compatible
;; OPERATING SYSTEM: Windows 95
;; SOFTWARE: FastSEQ for Windows Version 2.0b
;; CURRENT APPLICATION DATA:
;; APPLICATION NUMBER: US/09/144,925
;; FILING DATE:
;; CLASSIFICATION:
;; PRIOR APPLICATION DATA:
;; APPLICATION NUMBER: 08/685,992
;; FILING DATE: July 25, 1996
;; ATTORNEY/AGENT INFORMATION:
;; NAME: Granahan, Patricia
;; REGISTRATION NUMBER: 32,227
;; REFERENCE/DOCKET NUMBER: CSHL96-03Z
;; TELEPHONE: 781-861-6240
;; TELEFAX: 781-861-9540
;; TELEX:
;; INFORMATION FOR SEQ ID NO: 12:
;; SEQUENCE CHARACTERISTICS:
;; LENGTH: 266 amino acids

;
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: peptide
US-09-144-925-12

Query Match 1.4%; Score 7; DB 2; Length 266;
Best Local Similarity 100.0%; Pred. No. 81;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 473 IIREKGV 479
| | | | |
DB 215 IIREKGV 221

RESULT 7
US-08-286-819A-23
; Sequence 23, Application US/08286819A
; Patent No. 5871910
; GENERAL INFORMATION:

; APPLICANT: ARTHUR, MICHEL
; APPLICANT: DUKTA-MALEN, SYLVIE
; APPLICANT: MOLINAS, CATHERINE
; APPLICANT: COURVALIN, PATRICE
; TITLE OF INVENTION: POLYPEPTIDES IMPLICATED IN THE
; TITLE OF INVENTION: EXPRESSION OF RESISTANCE TO GLYCOPOLYMERES, IN PARTICULAR
; TITLE OF INVENTION: IN GRAM-POSITIVE BACTERIA, NUCLEOTIDE SEQUENCE CODING FOR
; TITLE OF INVENTION: THESE POLYPEPTIDES AND USE FOR DIAGNOSIS
; NUMBER OF SEQUENCES: 54
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: OBLON, SPIVAK, MCCLELLAND, MAIER & NEUSTADT,
; ADDRESS: 1755 S. Jefferson Davis Highway, Suite 400
; CITY: Arlington
; STATE: Virginia
; COUNTRY: U.S.A.
; ZIP: 22202

; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/286,819A
; FILING DATE: 05-AUG-1994
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/174,682
; FILING DATE: 28-DEC-1993
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/917,146
; FILING DATE: 10-AUG-1992
; CLASSIFICATION: 435

; APPLICATION NUMBER: PCT/FR/91/00855
; FILING DATE: 29-OCT-1991
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: FR 9013579
; FILING DATE: 31-OCT-1990
; CLASSIFICATION: 435
; NAME: Oblon, No. 5871910man F.
; REGISTRATION NUMBER: 24,618
; REFERENCE/DOCKET NUMBER: 660-060-0 PCT
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (703) 413-3000
; TELEFAX: (703) 413-2220
; TELEX: 248855 OPAT UR
; INFORMATION FOR SEQ ID NO: 23:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 303 amino acids
; TYPE: amino acid

;
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-08-286-819A-23

Query Match 1.4%; Score 7; DB 2; Length 303;
Best Local Similarity 100.0%; Pred. No. 91;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 215 KWIEENA 221
| | | | |
DB 185 KWIEENA 191

RESULT 8
US-08-980-357-23
; Sequence 23, Application US/08980357
; Patent No. 6013508
; GENERAL INFORMATION:

; APPLICANT: ARTHUR, MICHEL
; APPLICANT: DUKTA-MALEN, SYLVIE
; APPLICANT: MOLINAS, CATHERINE
; APPLICANT: COURVALIN, PATRICE
; TITLE OF INVENTION: POLYPEPTIDES IMPLICATED IN THE
; TITLE OF INVENTION: EXPRESSION OF RESISTANCE TO GLYCOPOLYMERES, IN PARTICULAR
; TITLE OF INVENTION: IN GRAM-POSITIVE BACTERIA, NUCLEOTIDE SEQUENCE CODING FOR
; TITLE OF INVENTION: THESE POLYPEPTIDES AND USE FOR DIAGNOSIS
; NUMBER OF SEQUENCES: 54
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: OBLON, SPIVAK, MCCLELLAND, MAIER & NEUSTADT,
; ADDRESS: 1755 S. Jefferson Davis Highway, Suite 400
; CITY: Arlington
; STATE: Virginia
; COUNTRY: U.S.A.
; ZIP: 22202

; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/980,357
; FILING DATE:
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/286,819
; FILING DATE: 05-AUG-1994
; APPLICATION NUMBER: US 08/174,682
; FILING DATE: 28-DEC-1993
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/917,146
; FILING DATE: 10-AUG-1992
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: PCT/FR/91/00855
; FILING DATE: 29-OCT-1991
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: FR 9013579
; FILING DATE: 31-OCT-1990
; ATTORNEY/AGENT INFORMATION:
; NAME: Oblon, No. 6013508man F.
; REGISTRATION NUMBER: 24,618
; REFERENCE/DOCKET NUMBER: 660-060-0 PCT
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (703) 413-3000
; TELEFAX: (703) 413-2220
; TELEX: 248855 OPAT UR
; INFORMATION FOR SEQ ID NO: 23:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 303 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-08-980-357-23

Query Match 1.4%; Score 7; DB 3; Length 303;
Best Local Similarity 100.0%; Pred. No. 91;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 215 KWIEENA 221
DB 185 KWIEENA 191

RESULT 9

US-08-481-377-6
; Sequence 6, Application US/08481377
; Patent No. 5808007
; GENERAL INFORMATION:
; APPLICANT: JOHNS HOPKINS UNIVERSITY
; TITLE OF INVENTION: GROWTH DIFFERENTIATION FACTOR-3
; NUMBER OF SEQUENCES: 29
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: SPENSLEY HORN JUBAS & LUBITZ
; STREET: 1880 CENTURY PARK EAST, FIFTH FLOOR
; CITY: LOS ANGELES
; STATE: CALIFORNIA
; COUNTRY: US
; ZIP: 90067
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/481.377
; FILING DATE:
; CLASSIFICATION: 514
; PRIOR APPLICATION NUMBER: PCT/US94/00666
; FILING DATE: 12-JAN-1994
; ATTORNEY/AGENT INFORMATION:
; NAME: WETHERELL, JR. Ph.D., JOHN R.
; REGISTRATION NUMBER: 31,678
; REFERENCE/DOCKET NUMBER: FD2279 PCT
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (619) 455-5100
; TELEFAX: (619) 455-5110
; INFORMATION FOR SEQ ID NO: 6:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 366 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-08-481-377-6

Query Match 1.4%; Score 7; DB 1; Length 366;
Best Local Similarity 100.0%; Pred. No. 1.1e+02;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 469 VLRKIIR 475
DB 51 VLRKIIR 57

RESULT 10

US-08-827-336-9
; Sequence 9, Application US/08827336
; Patent No. 6004780
; GENERAL INFORMATION:
; APPLICANT: SOPPET, DANIEL
; TITLE OF INVENTION: GROWTH FACTOR HTER36
; NUMBER OF SEQUENCES: 9
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: HUMAN GENOME SCIENCES, INC.
; STREET: 9410 KEY WEST AVENUE
; CITY: ROCKVILLE

STATE: MARYLAND
COUNTRY: USA
ZIP: 20850
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/827.336
FILING DATE:
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: BROOKES, ANDY, A.
REGISTRATION NUMBER: 36,373
REFERENCE/DOCKET NUMBER: PF230
TELECOMMUNICATION INFORMATION:
TELEPHONE: 301-309-8504
TELEFAX: 301-309-8512
INFORMATION FOR SEQ ID NO: 9:
SEQUENCE CHARACTERISTICS:
LENGTH: 366 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-827-336-9

Query Match 1.4%; Score 7; DB 3; Length 366;
Best Local Similarity 100.0%; Pred. No. 1.1e+02;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 469 VLRKIIR 475
DB 51 VLRKIIR 57

RESULT 11

US-09-153-733A-6
; Sequence 6, Application US/09153733A
; Patent No. 6025475
; GENERAL INFORMATION:
; APPLICANT: JOHNS HOPKINS UNIVERSITY
; TITLE OF INVENTION: GROWTH DIFFERENTIATION FACTOR-3
; NUMBER OF SEQUENCES: 29
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: SPENSLEY HORN JUBAS & LUBITZ
; STREET: 1880 CENTURY PARK EAST, FIFTH FLOOR
; CITY: LOS ANGELES
; STATE: CALIFORNIA
; COUNTRY: US
; ZIP: 90067
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/153.733A
; FILING DATE:
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/481.377
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: WETHERELL, JR. Ph.D., JOHN R.
; REGISTRATION NUMBER: 31,678
; REFERENCE/DOCKET NUMBER: FD2279 PCT
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (619) 455-5100
; TELEFAX: (619) 455-5110
; INFORMATION FOR SEQ ID NO: 6:
; SEQUENCE CHARACTERISTICS:

LENGTH: 366 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
US-09-153-733A-6

Query Match 1.4%; Score 7; DB 3; Length 366;
Best Local Similarity 100.0%; Pred. No. 1.1e+02;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 469 VLRKIIR 475
Db 51 VLRKIIR 57
|||||||

RESULT 12
US-09-389-705-6
Sequence 6, Application US/09389705
Patent No. 6391565
GENERAL INFORMATION:

APPLICANT: JOHNS HOPKINS UNIVERSITY
TITLE OF INVENTION: GROWTH DIFFERENTIATION FACTOR-3
NUMBER OF SEQUENCES: 29
CORRESPONDENCE ADDRESS:
ADDRESSEE: SPENSLEY HORN JUBAS & LUBITZ
STREET: 1880 CENTURY PARK EAST, FIFTH FLOOR
CITY: LOS ANGELES
STATE: CALIFORNIA
COUNTRY: US
ZIP: 90067

COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/389,705
FILING DATE: 03-Sep-1999
CLASSIFICATION: <Unknown>

PRIOR APPLICATION DATA:
APPLICATION NUMBER: 09/153,733
FILING DATE: <Unknown>
ATTORNEY/AGENT INFORMATION:
NAME: WETHERELL, JR. Ph.D., JOHN R.
REGISTRATION NUMBER: 31,678
REFERENCE/DOCKET NUMBER: FD2279 PCT

TELECOMMUNICATION INFORMATION:
TELEPHONE: (619) 455-5100
TELEFAX: (619) 455-5110
INFORMATION FOR SEQ ID NO: 6:
SEQUENCE CHARACTERISTICS:
LENGTH: 366 amino acids
TOPOLOGY: linear
MOLECULE TYPE: protein
SEQUENCE DESCRIPTION: SEQ ID NO: 6:
US-09-389-705-6

Query Match 1.4%; Score 7; DB 4; Length 366;
Best Local Similarity 100.0%; Pred. No. 1.1e+02;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 469 VLRKIIR 475
Db 51 VLRKIIR 57
|||||||

RESULT 13
US-09-357-905-9
Sequence 9, Application US/09357905
Patent No. 6413933
GENERAL INFORMATION:
APPLICANT: SOPPET, DANIEL

TITLE OF INVENTION: GROWTH FACTOR HTER36
NUMBER OF SEQUENCES: 9
CORRESPONDENCE ADDRESS:
ADDRESSEE: HUMAN GENOME SCIENCES, INC.
STREET: 9410 KEY WEST AVENUE
CITY: ROCKVILLE
STATE: MARYLAND
COUNTRY: USA
ZIP: 20850
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/357,905
FILING DATE: 21-Jul-1999
CLASSIFICATION: <Unknown>
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/827,336
FILING DATE: <Unknown>
ATTORNEY/AGENT INFORMATION:
NAME: BROOKES, ANDY, A.
REGISTRATION NUMBER: 36,373
REFERENCE/DOCKET NUMBER: PF230
TELECOMMUNICATION INFORMATION:
TELEPHONE: 301-309-8504
TELEFAX: 301-309-8512
INFORMATION FOR SEQ ID NO: 9:
SEQUENCE CHARACTERISTICS:
LENGTH: 366 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: protein
SEQUENCE DESCRIPTION: SEQ ID NO: 9:
US-09-357-905-9

Query Match 1.4%; Score 7; DB 4; Length 366;
Best Local Similarity 100.0%; Pred. No. 1.1e+02;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 469 VLRKIIR 475
Db 51 VLRKIIR 57
|||||||

RESULT 14
PCT-US94-00666-6
Sequence 6, Application PC/TUS9400666
GENERAL INFORMATION:
APPLICANT: JOHNS HOPKINS UNIVERSITY
TITLE OF INVENTION: GROWTH DIFFERENTIATION FACTOR-3
NUMBER OF SEQUENCES: 29
CORRESPONDENCE ADDRESS:
ADDRESSEE: SPENSLEY HORN JUBAS & LUBITZ
STREET: 1880 CENTURY PARK EAST, FIFTH FLOOR
CITY: LOS ANGELES
STATE: CALIFORNIA
COUNTRY: US
ZIP: 90067
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: PCT/US94/00666
FILING DATE: 12-JAN-1994
CLASSIFICATION:
ATTORNEY/AGENT INFORMATION:
NAME: WETHERELL, JR. Ph.D., JOHN R.
REGISTRATION NUMBER: 31,678

```

; REFERENCE/DOCKET NUMBER: FD2279 PCT
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (619) 455-5100
; TELEFAX: (619) 455-5110
; INFORMATION FOR SEQ ID NO: 6:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 366 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
PCT-US94-00666-6

```

```

Query Match      1.4%; Score 7; DB 5; Length 366;
Best Local Similarity 100.0%; Pred. No. 1.1e+02;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

```

QY 469 VLKRIIR 475
Db 51 VLKRIIR 57

```

```

RESULT 15
US-09-269-137-4
; Sequence 4, Application US/09269137
; Patent No. 6413751
; GENERAL INFORMATION:
; APPLICANT: Shapiro, Lucy
; APPLICANT: Kahng, Lyn Sue
; APPLICANT: Benkovic, Stephen J
; APPLICANT: Wright, Rachel
; APPLICANT: Stephens, Craig
; APPLICANT: Berdis, Anthony
; APPLICANT: Lee, Irene
; TITLE OF INVENTION: DNA Adenine Methyltransferases and Uses Thereof
; FILE REFERENCE: 99-292-B
; CURRENT APPLICATION NUMBER: US/09/269,137
; CURRENT FILING DATE: 1999-07-19
; NUMBER OF SEQ ID NOS: 24
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 4
; LENGTH: 377
; TYPE: PRT
; ORGANISM: Brucella abortus
US-09-269-137-4

```

```

Query Match      1.4%; Score 7; DB 4; Length 377;
Best Local Similarity 100.0%; Pred. No. 1.1e+02;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

```

QY 470 LRKIIRE 476
Db 364 LRKIIRE 370

```

```

Search completed: May 8, 2003, 16:35:15
Job time : 40 secs

```

THIS PAGE BLANK (UGPTD)

GenCore version 5.1.4_p5_4578
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: May 8, 2003, 16:33:44 ; Search time 32 Seconds
(without alignments)
1394.765 Million cell updates/sec

Title: US-09-926-163B-2

Perfect score: 485

Sequence: 1 MITRETKSLPANVOAPDY.....KVILRKIIREKGVKAAIPA 485

Scoring table: OLIGO

Gapop 60.0 , Gapext 60.0

Searched: 349150 seqs, 92025710 residues

Word size : 0

Total number of hits satisfying chosen parameters: 349150

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Database : Published Applications_AA:*

- 1: /cgn2_6/ptodata/1/pubpaa/US08_NEW_PUB.pep.*
- 2: /cgn2_6/ptodata/1/pubpaa/PCT_NEW_PUB.pep.*
- 3: /cgn2_6/ptodata/1/pubpaa/US06_NEW_PUB.pep.*
- 4: /cgn2_6/ptodata/1/pubpaa/US06_PUBCOMB.pep.*
- 5: /cgn2_6/ptodata/1/pubpaa/US07_NEW_PUB.pep.*
- 6: /cgn2_6/ptodata/1/pubpaa/US07_PUBCOMB.pep.*
- 7: /cgn2_6/ptodata/1/pubpaa/PCTUS_PUBCOMB.pep.*
- 8: /cgn2_6/ptodata/1/pubpaa/US08_PUBCOMB.pep.*
- 9: /cgn2_6/ptodata/1/pubpaa/US09_NEW_PUB.pep.*
- 10: /cgn2_6/ptodata/1/pubpaa/US09_PUBCOMB.pep.*
- 11: /cgn2_6/ptodata/1/pubpaa/US10_NEW_PUB.pep.*
- 12: /cgn2_6/ptodata/1/pubpaa/US10_PUBCOMB.pep.*
- 13: /cgn2_6/ptodata/1/pubpaa/US60_NEW_PUB.pep.*
- 14: /cgn2_6/ptodata/1/pubpaa/US60_PUBCOMB.pep.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	ID	Description
1	9	1.9	503	9 US-09-738-626-3626
2	8	1.6	163	9 US-10-028-072-504
3	8	1.6	163	9 US-10-121-049-504
4	8	1.6	163	9 US-10-123-904-504
5	8	1.6	163	9 US-10-140-470-504
6	8	1.6	163	9 US-10-175-746-504
7	8	1.6	163	9 US-10-176-918-504
8	8	1.6	163	9 US-10-176-921-504
9	8	1.6	163	9 US-10-227-884-184
10	8	1.6	163	9 US-10-137-865-504
11	8	1.6	163	9 US-10-140-474-504
12	8	1.6	163	9 US-10-142-431-504
13	8	1.6	163	9 US-10-143-114-504
14	8	1.6	163	9 US-10-230-163-184
15	8	1.6	163	9 US-10-140-002-504
16	8	1.6	163	9 US-10-006-856A-158
17	8	1.6	163	9 US-10-142-419-504
18	8	1.6	163	9 US-10-218-631-184
19	8	1.6	163	9 US-10-230-338-184

20	8	1.6	163	9	US-10-123-262-504	Sequence 504, App
21	8	1.6	163	9	US-10-142-423-504	Sequence 504, App
22	8	1.6	163	9	US-10-230-414-184	Sequence 184, App
23	8	1.6	163	9	US-10-006-818A-158	Sequence 158, App
24	8	1.6	163	9	US-10-121-050-504	Sequence 504, App
25	8	1.6	163	9	US-10-141-755-504	Sequence 504, App
26	8	1.6	163	9	US-10-143-032-504	Sequence 504, App
27	8	1.6	163	9	US-10-015-393A-158	Sequence 158, App
28	8	1.6	163	9	US-10-123-108-504	Sequence 504, App
29	8	1.6	163	9	US-10-123-236-504	Sequence 504, App
30	8	1.6	163	9	US-10-123-261-504	Sequence 504, App
31	8	1.6	163	9	US-10-140-921-504	Sequence 504, App
32	8	1.6	163	9	US-10-140-928-504	Sequence 504, App
33	8	1.6	163	9	US-10-216-159A-184	Sequence 184, App
34	8	1.6	163	9	US-09-946-374-158	Sequence 158, App
35	8	1.6	163	9	US-10-012-121A-158	Sequence 158, App
36	8	1.6	163	9	US-10-015-869A-158	Sequence 158, App
37	8	1.6	163	9	US-10-121-045-504	Sequence 504, App
38	8	1.6	163	9	US-10-123-292-504	Sequence 504, App
39	8	1.6	163	9	US-10-123-903-504	Sequence 504, App
40	8	1.6	163	9	US-10-124-819-504	Sequence 504, App
41	8	1.6	163	9	US-10-124-822-504	Sequence 504, App
42	8	1.6	163	9	US-10-140-925-504	Sequence 504, App
43	8	1.6	163	9	US-10-160-498-504	Sequence 504, App
44	8	1.6	163	9	US-10-218-849-184	Sequence 184, App
45	8	1.6	163	9	US-10-227-873-184	Sequence 184, App

ALIGNMENTS

RESULT 1
US-09-738-626-3626
; Sequence 3626, Application US/09738626
; Publication No. US20020197605A1
; GENERAL INFORMATION:
; APPLICANT: NAKAGAWA, SATOSHI
; APPLICANT: MIZOGUCHI, HIROSHI
; APPLICANT: ANDO, SEIKO
; APPLICANT: HAYASHI, MIKIRO
; APPLICANT: OCHIAI, KEIKO
; APPLICANT: YOKOI, HARUHIKO
; APPLICANT: TATEISHI, NAKOYO
; APPLICANT: SENOH, AKIHIRO
; APPLICANT: IKEDA, MASATO
; APPLICANT: OZAKI, AKIO
; TITLE OF INVENTION: NOVEL POLYNUCLEOTIDES
; FILE REFERENCE: 249-125
; CURRENT APPLICATION NUMBER: US/09/738,626
; CURRENT FILING DATE: 2000-12-18
; PRIOR APPLICATION NUMBER: JP 99/377484
; PRIOR FILING DATE: 1999-12-16
; PRIOR APPLICATION NUMBER: JP 00/159162
; PRIOR FILING DATE: 2000-04-07
; PRIOR APPLICATION NUMBER: JP 00/280988
; PRIOR FILING DATE: 2000-08-03
; NUMBER OF SEQ ID NOS: 7059
; SOFTWARE: PatentIn ver. 3.0
; SEQ ID NO 3626
; LENGTH: 503
; TYPE: PRT
; ORGANISM: Corynebacterium glutamicum
US-09-738-626-3626

Query Match 1.9%; Score 9; DB 9; Length 503;
Best Local Similarity 100.0%; Pred. No. 2.2;
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 26 PGIVHFGVG 34

Db 33 PGIVHFGVG 41

RESULT 2
US-10-028-072-504
: Sequence 504, Application US/10028072
: Publication No. US20030004311A1
: GENERAL INFORMATION:
: APPLICANT: Baker, Kevin P.
: APPLICANT: Beresini, Maureen
: APPLICANT: DeForge, Laura
: APPLICANT: Desnoyers, Luc
: APPLICANT: Filvaroff, Ellen
: APPLICANT: Gao, Wei-Qiang
: APPLICANT: Gerritsen, Mary E.
: APPLICANT: Goddard, Audrey
: APPLICANT: Godowski, Paul J.
: APPLICANT: Gurney, Austin L.
: APPLICANT: Sherwood, Steven
: APPLICANT: Smith, Victoria
: APPLICANT: Stewart, Timothy A.
: APPLICANT: Tumas, Daniel
: APPLICANT: Watanabe, Colin K
: APPLICANT: Wood, William
: APPLICANT: Zhang
: TITLE OF INVENTION:
: FILE REFERENCE:
: CURRENT APPLICATION NUMBER: US/10/028,072
: CURRENT FILING DATE: 2001-12-19
: PRIOR APPLICATION NUMBER: 60/049911
: PRIOR FILING DATE: 1997-06-18
: PRIOR APPLICATION NUMBER: 60/056974
: PRIOR FILING DATE: 1997-08-26
: PRIOR APPLICATION NUMBER: 60/059113
: PRIOR FILING DATE: 1997-09-17
: PRIOR APPLICATION NUMBER: 60/059115
: PRIOR FILING DATE: 1997-09-17
: PRIOR APPLICATION NUMBER: 60/059117
: PRIOR FILING DATE: 1997-09-17
: PRIOR APPLICATION NUMBER: 60/059122
: PRIOR FILING DATE: 1997-09-17
: PRIOR APPLICATION NUMBER: 60/059184
: PRIOR FILING DATE: 1997-09-17
: PRIOR APPLICATION NUMBER: 60/059263
: PRIOR FILING DATE: 1997-09-18
: PRIOR APPLICATION NUMBER: 60/059352
: PRIOR FILING DATE: 1997-09-19
: PRIOR APPLICATION NUMBER: 60/059588
: PRIOR FILING DATE: 1997-09-19
: PRIOR APPLICATION NUMBER: 60/059836
: PRIOR FILING DATE: 1997-09-24
: PRIOR APPLICATION NUMBER: 60/062250
: PRIOR FILING DATE: 1997-10-17
: PRIOR APPLICATION NUMBER: 60/062285
: PRIOR FILING DATE: 1997-10-17
: PRIOR APPLICATION NUMBER: 60/062287
: PRIOR FILING DATE: 1997-10-17
: PRIOR APPLICATION NUMBER: 60/062814
: PRIOR FILING DATE: 1997-10-24
: PRIOR APPLICATION NUMBER: 60/062816
: PRIOR FILING DATE: 1997-10-24
: PRIOR APPLICATION NUMBER: 60/063045
: PRIOR FILING DATE: 1997-10-24
: PRIOR APPLICATION NUMBER: 60/063082
: PRIOR FILING DATE: 1997-10-31
: PRIOR APPLICATION NUMBER: 60/063127
: PRIOR FILING DATE: 1997-10-24
: PRIOR APPLICATION NUMBER: 60/063327
: PRIOR FILING DATE: 1997-10-27
: PRIOR APPLICATION NUMBER: 60/063329
: PRIOR FILING DATE: 1997-10-27
: PRIOR APPLICATION NUMBER: 60/063550
: PRIOR FILING DATE: 1997-10-28
: PRIOR APPLICATION NUMBER: 60/063561
: PRIOR FILING DATE: 1997-10-28
: PRIOR APPLICATION NUMBER: 60/063704
: PRIOR FILING DATE: 1997-10-29
: PRIOR APPLICATION NUMBER: 60/063733
: PRIOR FILING DATE: 1997-10-29
: PRIOR APPLICATION NUMBER: 60/063735
: PRIOR FILING DATE: 1997-10-29
: PRIOR APPLICATION NUMBER: 60/063738
: PRIOR FILING DATE: 1997-10-29
: PRIOR APPLICATION NUMBER: 60/063755
: PRIOR FILING DATE: 1997-10-17
: PRIOR APPLICATION NUMBER: 60/064248
: PRIOR FILING DATE: 1997-11-03
: PRIOR APPLICATION NUMBER: 60/064809
: PRIOR FILING DATE: 1997-11-07
: PRIOR APPLICATION NUMBER: 60/065186
: PRIOR FILING DATE: 1997-11-12
: PRIOR APPLICATION NUMBER: 60/065846
: PRIOR FILING DATE: 1997-11-17
: PRIOR APPLICATION NUMBER: 60/066364
: PRIOR FILING DATE: 1997-11-21
: PRIOR APPLICATION NUMBER: 60/066453
: PRIOR FILING DATE: 1997-11-24
: PRIOR APPLICATION NUMBER: 60/066511
: PRIOR FILING DATE: 1997-11-24
: PRIOR APPLICATION NUMBER: 60/066770
: PRIOR FILING DATE: 1997-11-24
: PRIOR APPLICATION NUMBER: 60/069212
: PRIOR FILING DATE: 1997-12-11
: PRIOR APPLICATION NUMBER: 60/069278
: PRIOR FILING DATE: 1997-12-11
: PRIOR APPLICATION NUMBER: 60/069334
: PRIOR FILING DATE: 1997-12-11
: PRIOR APPLICATION NUMBER: 60/069694
: PRIOR FILING DATE: 1997-12-16
: PRIOR APPLICATION NUMBER: 60/072320
: PRIOR FILING DATE: 1998-01-23
: PRIOR APPLICATION NUMBER: 60/073612
: PRIOR FILING DATE: 1998-02-04
: PRIOR APPLICATION NUMBER: 60/074086
: PRIOR FILING DATE: 1998-02-09
: PRIOR APPLICATION NUMBER: 60/074092
: PRIOR FILING DATE: 1998-02-09
: PRIOR APPLICATION NUMBER: 60/077791
: PRIOR FILING DATE: 1998-03-12
: PRIOR APPLICATION NUMBER: 60/078910
: PRIOR FILING DATE: 1998-03-20
: PRIOR APPLICATION NUMBER: 60/079294
: PRIOR FILING DATE: 1998-03-25
: PRIOR APPLICATION NUMBER: 60/079663
: PRIOR FILING DATE: 1998-02-27
: PRIOR APPLICATION NUMBER: 60/079728
: PRIOR FILING DATE: 1998-03-27
: PRIOR APPLICATION NUMBER: 60/080165
: PRIOR FILING DATE: 1998-03-31
: PRIOR APPLICATION NUMBER: 60/081203
: PRIOR FILING DATE: 1998-04-09
: PRIOR APPLICATION NUMBER: 60/081229
: PRIOR FILING DATE: 1998-04-09
: PRIOR APPLICATION NUMBER: 60/081695
: PRIOR FILING DATE: 1998-04-14
: PRIOR APPLICATION NUMBER: 60/081817
: PRIOR FILING DATE: 1998-04-15
: PRIOR APPLICATION NUMBER: 60/081818
: PRIOR FILING DATE: 1998-04-15
: PRIOR APPLICATION NUMBER: 60/082999
: PRIOR FILING DATE: 1998-04-24
: PRIOR APPLICATION NUMBER: 60/083322
: PRIOR FILING DATE: 1998-04-28
: PRIOR APPLICATION NUMBER: 60/083545
: PRIOR FILING DATE: 1998-04-29
: PRIOR APPLICATION NUMBER: 60/084600
: PRIOR FILING DATE: 1998-05-07
: PRIOR APPLICATION NUMBER: 60/084627
: PRIOR FILING DATE: 1998-05-07

;; PRIOR APPLICATION NUMBER: 60/084637
;; PRIOR FILING DATE: 1998-05-07
;; PRIOR APPLICATION NUMBER: 60/085149
;; PRIOR FILING DATE: 1998-05-12
;; PRIOR APPLICATION NUMBER: 60/085323
;; PRIOR FILING DATE: 1998-05-13
;; PRIOR APPLICATION NUMBER: 60/085338
;; PRIOR FILING DATE: 1998-05-13
;; PRIOR APPLICATION NUMBER: 60/085339
;; PRIOR FILING DATE: 1998-05-13
;; PRIOR APPLICATION NUMBER: 60/085579
;; PRIOR FILING DATE: 1998-05-15
;; PRIOR APPLICATION NUMBER: 60/085697
;; PRIOR FILING DATE: 1998-05-15
;; PRIOR APPLICATION NUMBER: 60/085704
;; PRIOR FILING DATE: 1998-05-15
;; PRIOR APPLICATION NUMBER: 60/086414
;; PRIOR FILING DATE: 1998-05-22
;; PRIOR APPLICATION NUMBER: 60/086430
;; PRIOR FILING DATE: 1998-05-22
;; PRIOR APPLICATION NUMBER: 60/087106
;; PRIOR FILING DATE: 1998-05-28
;; PRIOR APPLICATION NUMBER: 60/088026
;; PRIOR FILING DATE: 1998-06-04
;; PRIOR APPLICATION NUMBER: 60/088730
;; PRIOR FILING DATE: 1998-06-10
;; PRIOR APPLICATION NUMBER: 60/088741
;; PRIOR FILING DATE: 1998-06-10
;; PRIOR APPLICATION NUMBER: 60/088810
;; PRIOR FILING DATE: 1998-06-10
;; PRIOR APPLICATION NUMBER: 60/088858
;; PRIOR FILING DATE: 1998-06-11
;; PRIOR APPLICATION NUMBER: 60/089532
;; PRIOR FILING DATE: 1998-06-17
;; PRIOR APPLICATION NUMBER: 60/089599
;; PRIOR FILING DATE: 1998-06-17
;; PRIOR APPLICATION NUMBER: 60/089907
;; PRIOR FILING DATE: 1998-06-18
;; PRIOR APPLICATION NUMBER: 60/089947
;; PRIOR FILING DATE: 1998-06-19
;; PRIOR APPLICATION NUMBER: 60/090349
;; PRIOR FILING DATE: 1998-06-23
;; PRIOR APPLICATION NUMBER: 60/090429
;; PRIOR FILING DATE: 1998-06-24
;; PRIOR APPLICATION NUMBER: 60/090445
;; PRIOR FILING DATE: 1998-06-24
;; PRIOR APPLICATION NUMBER: 60/090538
;; PRIOR FILING DATE: 1998-06-24
;; PRIOR APPLICATION NUMBER: 60/090863
;; PRIOR FILING DATE: 1998-06-26
;; PRIOR APPLICATION NUMBER: 60/091360
;; PRIOR FILING DATE: 1998-07-01
;; PRIOR APPLICATION NUMBER: 60/091519
;; PRIOR FILING DATE: 1998-07-02
;; PRIOR APPLICATION NUMBER: 60/091982
;; PRIOR FILING DATE: 1998-07-07

Query Match 1.6%; Score 8; DB 9; Length 163;
Best Local Similarity 100.0%; Pred. No. 7.9;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 81 LYSLTETA 88
|||||
Db 136 LYSLTETA 143

RESULT 3

US-10-121-049-504
; Sequence 504, Application US/10121049
; Publication No. US20030022239A1
; GENERAL INFORMATION:
; APPLICANT: Baker, Kevin P.
; APPLICANT: Beresini, Maureen

;; APPLICANT: DeForge, Laura
;; APPLICANT: Desnoyers, Luc
;; APPLICANT: Filvaroff, Ellen
;; APPLICANT: Gao, Wei-Qiang
;; APPLICANT: Gerritsen, Mary E.
;; APPLICANT: Goddard, Audrey
;; APPLICANT: Godowski, Paul J.
;; APPLICANT: Gurney, Austin L.
;; APPLICANT: Sherwood, Steven
;; APPLICANT: Smith, Victoria
;; APPLICANT: Stewart, Timothy A.
;; APPLICANT: Tumas, Daniel
;; APPLICANT: Watanabe, Colin K
;; APPLICANT: Wood, William
;; APPLICANT: Zhang, Zemin
;; TITLE OF INVENTION: SECRETED AND TRANSMEMBRANE POLYPEPTIDES AND NUCLEIC
;; FILE REFERENCE: P3330R1C17
;; CURRENT APPLICATION NUMBER: US/10/121,049
;; CURRENT FILING DATE: 2002-04-12
;; Prior Application removed - See File Wrapper or Palm
;; NUMBER OF SEQ ID NOS: 550
;; SEQ ID NO 504
;; LENGTH: 163
;; TYPE: PRT
;; ORGANISM: Homo Sapien
US-10-121-049-504

Query Match 1.6%; Score 8; DB 9; Length 163;
Best Local Similarity 100.0%; Pred. No. 7.9;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 81 LYSLTETA 88
|||||
Db 136 LYSLTETA 143

RESULT 4

US-10-123-904-504
; Sequence 504, Application US/10123904
; Publication No. US20030022328A1
; GENERAL INFORMATION:

;; APPLICANT: Baker, Kevin P.
;; APPLICANT: Beresini, Maureen
;; APPLICANT: DeForge, Laura
;; APPLICANT: Desnoyers, Luc
;; APPLICANT: Filvaroff, Ellen
;; APPLICANT: Gao, Wei-Qiang
;; APPLICANT: Gerritsen, Mary E.
;; APPLICANT: Goddard, Audrey
;; APPLICANT: Godowski, Paul J.
;; APPLICANT: Gurney, Austin L.
;; APPLICANT: Sherwood, Steven
;; APPLICANT: Smith, Victoria
;; APPLICANT: Stewart, Timothy A.
;; APPLICANT: Tumas, Daniel
;; APPLICANT: Watanabe, Colin K
;; APPLICANT: Wood, William
;; APPLICANT: Zhang, Zemin
;; TITLE OF INVENTION: SECRETED AND TRANSMEMBRANE POLYPEPTIDES AND NUCLEIC
;; FILE REFERENCE: P3330R1C54
;; CURRENT APPLICATION NUMBER: US/10/123,904
;; CURRENT FILING DATE: 2002-04-16
;; Prior Application removed - See File Wrapper or Palm
;; NUMBER OF SEQ ID NOS: 550
;; SEQ ID NO 504
;; LENGTH: 163
;; TYPE: PRT
;; ORGANISM: Homo Sapien
US-10-123-904-504

Query Match

1.6%; Score 8; DB 9; Length 163;

```
Best Local Similarity 100.0%; Pred. No. 7.9;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 81 LYSLTETA 88
Db 136 LYSLTETA 143

RESULT 5
US-10-140-470-504
; Sequence 504, Application US/10140470
; Publication No. US20030022331A1
; GENERAL INFORMATION:
; APPLICANT: Baker, Kevin P.
; APPLICANT: Beresini, Maureen
; APPLICANT: DeForge, Laura
; APPLICANT: Desnoyers, Luc
; APPLICANT: Filvaroff, Ellen
; APPLICANT: Gao, Wei-Qiang
; APPLICANT: Gerritsen, Mary E.
; APPLICANT: Goddard, Audrey
; APPLICANT: Godowski, Paul J.
; APPLICANT: Gurney, Austin L.
; APPLICANT: Sherwood, Steven
; APPLICANT: Smith, Victoria
; APPLICANT: Stewart, Timothy A.
; APPLICANT: Tumas, Daniel
; APPLICANT: Watanabe, Colin K
; APPLICANT: Wood, William
; APPLICANT: Zhang, Zemin
; TITLE OF INVENTION: SECRETED AND TRANSMEMBRANE POLYPEPTIDES AND NUCLEIC
; FILE REFERENCE: P333081C160
; CURRENT APPLICATION NUMBER: US/10/140,470
; CURRENT FILING DATE: 2002-05-06
; Prior Application removed - See Palm or File Wrapper
; NUMBER OF SEQ ID NOS: 550
; SEQ ID NO 504
; LENGTH: 163
; TYPE: PRT
; ORGANISM: Homo Sapien
US-10-140-470-504

Query Match 1.6%; Score 8; DB 9; Length 163;
Best Local Similarity 100.0%; Pred. No. 7.9;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 81 LYSLTETA 88
Db 136 LYSLTETA 143

RESULT 6
US-10-175-746-504
; Sequence 504, Application US/10175746
; Publication No. US20030027270A1
; GENERAL INFORMATION:
; APPLICANT: Baker, Kevin P.
; APPLICANT: Beresini, Maureen
; APPLICANT: DeForge, Laura
; APPLICANT: Desnoyers, Luc
; APPLICANT: Filvaroff, Ellen
; APPLICANT: Gao, Wei-Qiang
; APPLICANT: Gerritsen, Mary E.
; APPLICANT: Goddard, Audrey
; APPLICANT: Godowski, Paul J.
; APPLICANT: Gurney, Austin L.
; APPLICANT: Sherwood, Steven
; APPLICANT: Smith, Victoria
; APPLICANT: Stewart, Timothy A.
; APPLICANT: Tumas, Daniel
; APPLICANT: Watanabe, Colin K
; APPLICANT: Wood, William
; APPLICANT: Zhang, Zemin
; TITLE OF INVENTION: SECRETED AND TRANSMEMBRANE POLYPEPTIDES AND NUCLEIC
; FILE REFERENCE: P333081C382
; CURRENT APPLICATION NUMBER: US/10/176,918
; CURRENT FILING DATE: 2002-06-20
; Prior Application removed - See File Wrapper or Palm
; NUMBER OF SEQ ID NOS: 550
; SEQ ID NO 504
; LENGTH: 163
; TYPE: PRT
; ORGANISM: Homo Sapien
US-10-175-746-504

Query Match 1.6%; Score 8; DB 9; Length 163;
Best Local Similarity 100.0%; Pred. No. 7.9;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 81 LYSLTETA 88
Db 136 LYSLTETA 143

RESULT 7
US-10-176-918-504
; Sequence 504, Application US/10176918
; Publication No. US20030027275A1
; GENERAL INFORMATION:
; APPLICANT: Baker, Kevin P.
; APPLICANT: Beresini, Maureen
; APPLICANT: DeForge, Laura
; APPLICANT: Desnoyers, Luc
; APPLICANT: Filvaroff, Ellen
; APPLICANT: Gao, Wei-Qiang
; APPLICANT: Gerritsen, Mary E.
; APPLICANT: Goddard, Audrey
; APPLICANT: Godowski, Paul J.
; APPLICANT: Gurney, Austin L.
; APPLICANT: Sherwood, Steven
; APPLICANT: Smith, Victoria
; APPLICANT: Stewart, Timothy A.
; APPLICANT: Tumas, Daniel
; APPLICANT: Watanabe, Colin K
; APPLICANT: Wood, William
; APPLICANT: Zhang, Zemin
; TITLE OF INVENTION: SECRETED AND TRANSMEMBRANE POLYPEPTIDES AND NUCLEIC
; FILE REFERENCE: P333081C382
; CURRENT APPLICATION NUMBER: US/10/176,918
; CURRENT FILING DATE: 2002-06-20
; Prior Application removed - See File Wrapper or Palm
; NUMBER OF SEQ ID NOS: 550
; SEQ ID NO 504
; LENGTH: 163
; TYPE: PRT
; ORGANISM: Homo Sapien
US-10-176-918-504

Query Match 1.6%; Score 8; DB 9; Length 163;
Best Local Similarity 100.0%; Pred. No. 7.9;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 81 LYSLTETA 88
Db 136 LYSLTETA 143

RESULT 8
US-10-176-921-504
; Sequence 504, Application US/10176921
; Publication No. US20030027276A1
; GENERAL INFORMATION:
; APPLICANT: Baker, Kevin P.
```



```

; APPLICANT: Beresini, Maureen
; APPLICANT: Desforge, Laura
; APPLICANT: Desnoyers, Luc
; APPLICANT: Filvaroff, Ellen
; APPLICANT: Gao, Wei-Qiang
; APPLICANT: Gerritsen, Mary E.
; APPLICANT: Goddard, Audrey
; APPLICANT: Goddard, Audrey
; APPLICANT: Godowski, Paul J.
; APPLICANT: Gurney, Austin L.
; APPLICANT: Gurney, Austin L.
; APPLICANT: Sherwood, Steven
; APPLICANT: Smith, Victoria
; APPLICANT: Stewart, Timothy A.
; APPLICANT: Tamas, Daniel
; APPLICANT: Tamas, Daniel
; APPLICANT: Watanabe, Colin K
; APPLICANT: Wood, William
; APPLICANT: Zhang, Zemin
; TITLE OF INVENTION: SECRETED AND TRANSMEMBRANE POLYPEPTIDES AND NUCLEIC
; FILE REFERENCE: P3330RIC288
; CURRENT APPLICATION NUMBER: US/10/176,921
; CURRENT FILING DATE: 2002-06-20
; Prior Application removed - See File Wrapper or Palm
; NUMBER OF SEQ ID NOS: 550
; SEQ ID NO 504
; LENGTH: 163
; TYPE: PRT
; ORGANISM: Homo Sapien
; US-10-176-921-504

Query Match 1.6%; Score 8; DB 9; Length 163;
Best Local Similarity 100.0%; Pred. No. 7.9;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 81 LYSLTETA 88
Db 136 LYSLTETA 143

RESULT 9
US-10-227-884-184
; Sequence 184, Application US/10227884
; Publication No. US20030027988A1
; GENERAL INFORMATION:
; APPLICANT: Baker, Kevin P.
; APPLICANT: Desnoyers, Luc
; APPLICANT: Gerritsen, Mary
; APPLICANT: Goddard, Audrey
; APPLICANT: Godowski, Paul J.
; APPLICANT: Grimaldi, J. Christopher
; APPLICANT: Gurney, Austin L.
; APPLICANT: Smith, Victoria
; APPLICANT: Stephan, Jean-Philippe F.
; APPLICANT: Watanabe, Colin L.
; APPLICANT: Wood, William I.
; TITLE OF INVENTION: SECRETED AND TRANSMEMBRANE POLYPEPTIDES AND NUCLEIC
; FILE REFERENCE: P3330PIC79
; CURRENT APPLICATION NUMBER: US/10/227,884
; CURRENT FILING DATE: 2002-08-26
; PRIOR APPLICATION NUMBER: 10/119,480
; PRIOR FILING DATE: 2002-04-09
; PRIOR APPLICATION NUMBER: 60/059113
; PRIOR FILING DATE: 1997-09-17
; PRIOR APPLICATION NUMBER: 60/062287
; PRIOR FILING DATE: 1997-10-17
; PRIOR APPLICATION NUMBER: 60/063549
; PRIOR FILING DATE: 1997-10-28
; PRIOR APPLICATION NUMBER: 60/064103
; PRIOR FILING DATE: 1997-10-31
; PRIOR APPLICATION NUMBER: 60/069873
; PRIOR FILING DATE: 1997-12-17
; PRIOR APPLICATION NUMBER: 60/078910
; PRIOR FILING DATE: 1998-03-20
; PRIOR APPLICATION NUMBER: 60/079294
; PRIOR FILING DATE: 1998-03-25
; PRIOR APPLICATION NUMBER: 60/079656
; PRIOR FILING DATE: 1998-03-26
; PRIOR APPLICATION NUMBER: 60/079728
; PRIOR FILING DATE: 1998-03-27
; PRIOR APPLICATION NUMBER: 60/081819
; PRIOR FILING DATE: 1998-04-15
; PRIOR APPLICATION NUMBER: 60/081955
; PRIOR FILING DATE: 1998-04-15
; PRIOR APPLICATION NUMBER: 60/082804
; PRIOR FILING DATE: 1998-04-22
; PRIOR APPLICATION NUMBER: 60/084441
; PRIOR FILING DATE: 1998-05-06
; PRIOR APPLICATION NUMBER: 60/085323
; PRIOR FILING DATE: 1998-05-13
; PRIOR APPLICATION NUMBER: 60/085579
; PRIOR FILING DATE: 1998-05-15
; PRIOR APPLICATION NUMBER: 60/086392
; PRIOR FILING DATE: 1998-05-22
; PRIOR APPLICATION NUMBER: 60/089532
; PRIOR FILING DATE: 1998-06-17
; PRIOR APPLICATION NUMBER: 60/089538
; PRIOR FILING DATE: 1998-06-17
; PRIOR APPLICATION NUMBER: 60/089905
; PRIOR FILING DATE: 1998-06-18
; PRIOR APPLICATION NUMBER: 60/090472
; PRIOR FILING DATE: 1998-06-24
; PRIOR APPLICATION NUMBER: 60/090557
; PRIOR FILING DATE: 1998-06-24
; PRIOR APPLICATION NUMBER: 60/090691
; PRIOR FILING DATE: 1998-06-25
; PRIOR APPLICATION NUMBER: 60/090695
; PRIOR FILING DATE: 1998-06-25
; PRIOR APPLICATION NUMBER: 60/091982
; PRIOR FILING DATE: 1998-07-07
; PRIOR APPLICATION NUMBER: 60/095302
; PRIOR FILING DATE: 1998-08-04
; PRIOR APPLICATION NUMBER: 60/095318
; PRIOR FILING DATE: 1998-08-04
; PRIOR APPLICATION NUMBER: 60/095916
; PRIOR FILING DATE: 1998-08-10
; PRIOR APPLICATION NUMBER: 60/096146
; PRIOR FILING DATE: 1998-08-11
; PRIOR APPLICATION NUMBER: 60/096791
; PRIOR FILING DATE: 1998-08-17
; PRIOR APPLICATION NUMBER: 60/097986
; PRIOR FILING DATE: 1998-08-26
; PRIOR APPLICATION NUMBER: 60/098544
; PRIOR FILING DATE: 1998-08-31
; PRIOR APPLICATION NUMBER: 60/099596
; PRIOR FILING DATE: 1998-09-09
; PRIOR APPLICATION NUMBER: 60/099598
; PRIOR FILING DATE: 1998-09-09
; PRIOR APPLICATION NUMBER: 60/099803
; PRIOR FILING DATE: 1998-09-10
; PRIOR APPLICATION NUMBER: 60/099811
; PRIOR FILING DATE: 1998-09-10
; PRIOR APPLICATION NUMBER: 60/099812
; PRIOR FILING DATE: 1998-09-10
; PRIOR APPLICATION NUMBER: 60/099816
; PRIOR FILING DATE: 1998-09-10
; PRIOR APPLICATION NUMBER: 60/100038
; PRIOR FILING DATE: 1998-09-11
; PRIOR APPLICATION NUMBER: 60/100385
; PRIOR FILING DATE: 1998-09-15
; PRIOR APPLICATION NUMBER: 60/100390
; PRIOR FILING DATE: 1998-09-15
; PRIOR APPLICATION NUMBER: 60/100627
; PRIOR FILING DATE: 1998-09-16
; PRIOR APPLICATION NUMBER: 60/100848
; PRIOR FILING DATE: 1998-09-18
; PRIOR APPLICATION NUMBER: 60/100919
```

; PRIOR FILING DATE: 1998-09-17
 ; PRIOR APPLICATION NUMBER: 60/101477
 ; PRIOR FILING DATE: 1998-09-23
 ; PRIOR APPLICATION NUMBER: 60/101738
 ; PRIOR FILING DATE: 1998-09-24
 ; PRIOR APPLICATION NUMBER: 60/101741
 ; PRIOR FILING DATE: 1998-09-24
 ; PRIOR APPLICATION NUMBER: 60/101786
 ; PRIOR FILING DATE: 1998-09-25
 ; PRIOR APPLICATION NUMBER: 60/101916
 ; PRIOR FILING DATE: 1998-09-24
 ; PRIOR APPLICATION NUMBER: 60/101922
 ; PRIOR FILING DATE: 1998-09-24
 ; PRIOR APPLICATION NUMBER: 60/106178
 ; PRIOR FILING DATE: 1998-10-28
 ; PRIOR APPLICATION NUMBER: 60/106248
 ; PRIOR FILING DATE: 1998-10-29
 ; PRIOR APPLICATION NUMBER: 60/106464
 ; PRIOR FILING DATE: 1998-10-30
 ; PRIOR APPLICATION NUMBER: 60/106905
 ; PRIOR FILING DATE: 1998-11-03
 ; PRIOR APPLICATION NUMBER: 60/108787
 ; PRIOR FILING DATE: 1998-11-17
 ; PRIOR APPLICATION NUMBER: 60/108801
 ; PRIOR FILING DATE: 1998-11-17
 ; PRIOR APPLICATION NUMBER: 60/108849
 ; PRIOR FILING DATE: 1998-11-18
 ; PRIOR APPLICATION NUMBER: 60/112422
 ; PRIOR FILING DATE: 1998-12-15
 ; PRIOR APPLICATION NUMBER: 60/113296
 ; PRIOR FILING DATE: 1998-12-22
 ; PRIOR APPLICATION NUMBER: 60/113605
 ; PRIOR FILING DATE: 1998-12-23
 ; PRIOR APPLICATION NUMBER: 60/113621
 ; PRIOR FILING DATE: 1998-12-23
 ; PRIOR APPLICATION NUMBER: 60/115558
 ; PRIOR FILING DATE: 1999-01-12
 ; PRIOR APPLICATION NUMBER: 60/115565
 ; PRIOR FILING DATE: 1999-01-12
 ; PRIOR APPLICATION NUMBER: 60/115733
 ; PRIOR FILING DATE: 1999-01-12
 ; PRIOR APPLICATION NUMBER: 60/119549
 ; PRIOR FILING DATE: 1999-02-10
 ; PRIOR APPLICATION NUMBER: 60/123618
 ; PRIOR FILING DATE: 1999-03-10
 ; PRIOR APPLICATION NUMBER: 60/125259
 ; PRIOR FILING DATE: 1999-03-19
 ; PRIOR APPLICATION NUMBER: 60/125775
 ; PRIOR FILING DATE: 1999-03-23
 ; PRIOR APPLICATION NUMBER: 60/126773
 ; PRIOR FILING DATE: 1999-03-29
 ; PRIOR APPLICATION NUMBER: 60/127887
 ; PRIOR FILING DATE: 1999-04-05
 ; PRIOR APPLICATION NUMBER: 60/130232
 ; PRIOR FILING DATE: 1999-04-21
 ; PRIOR APPLICATION NUMBER: 60/131022
 ; PRIOR FILING DATE: 1999-04-26
 ; PRIOR APPLICATION NUMBER: 60/131270
 ; PRIOR FILING DATE: 1999-04-27
 ; PRIOR APPLICATION NUMBER: 60/131291
 ; PRIOR FILING DATE: 1999-04-27
 ; PRIOR APPLICATION NUMBER: 60/131445
 ; PRIOR FILING DATE: 1999-04-28
 ; PRIOR APPLICATION NUMBER: 60/134287
 ; PRIOR FILING DATE: 1999-05-14
 ; PRIOR APPLICATION NUMBER: 60/140650
 ; PRIOR FILING DATE: 1999-06-22
 ; PRIOR APPLICATION NUMBER: 60/140723
 ; PRIOR FILING DATE: 1999-06-22
 ; PRIOR APPLICATION NUMBER: 60/141037
 ; PRIOR FILING DATE: 1999-06-23
 ; PRIOR APPLICATION NUMBER: 60/144758
 ; PRIOR FILING DATE: 1999-07-20

; PRIOR APPLICATION NUMBER: 60/145698
 ; PRIOR FILING DATE: 1999-07-26
 ; PRIOR APPLICATION NUMBER: 60/146222
 ; PRIOR FILING DATE: 1999-07-28
 ; PRIOR APPLICATION NUMBER: 60/146963
 ; PRIOR FILING DATE: 1999-08-03
 ; PRIOR APPLICATION NUMBER: 60/149320
 ; PRIOR FILING DATE: 1999-08-17
 ; PRIOR APPLICATION NUMBER: 60/149638
 ; PRIOR FILING DATE: 1999-08-17
 ; PRIOR APPLICATION NUMBER: 60/151733
 ; PRIOR FILING DATE: 1999-08-31
 ; PRIOR APPLICATION NUMBER: 60/164418
 ; PRIOR FILING DATE: 1999-11-09
 ; PRIOR APPLICATION NUMBER: 60/166361
 ; PRIOR FILING DATE: 1999-11-16
 ; PRIOR APPLICATION NUMBER: 60/169445
 ; PRIOR FILING DATE: 1999-12-07
 ; PRIOR APPLICATION NUMBER: 60/169495
 ; PRIOR FILING DATE: 1999-12-07
 ; PRIOR APPLICATION NUMBER: 60/169835

Query Match 1.6%; Score 8; DB 9; Length 163;
 Best Local Similarity 100.0%; Pred.No. 7.9;
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 81 LYSUTETA 88
 |||||
 Db 136 LYSUTETA 143

RESULT 10

US-10-137-865-504
 ; Sequence 504, Application US/10137865.
 ; Publication No. US20030032155A1
 ; GENERAL INFORMATION:
 ; APPLICANT: Baker, Kevin P.
 ; APPLICANT: Beresini, Maureen
 ; APPLICANT: DeForge, Laura
 ; APPLICANT: Desnoyers, Luc
 ; APPLICANT: Filvaroff, Ellen
 ; APPLICANT: Gao, Wei-Qiang
 ; APPLICANT: Gerritsen, Mary E.
 ; APPLICANT: Goddard, Audrey
 ; APPLICANT: Godowski, Paul J.
 ; APPLICANT: Gurney, Austin L.
 ; APPLICANT: Sherwood, Steven
 ; APPLICANT: Smith, Victoria
 ; APPLICANT: Stewart, Timothy A.
 ; APPLICANT: Tamas, Daniel
 ; APPLICANT: Watanabe, Colin K
 ; APPLICANT: Wood, William
 ; APPLICANT: Zhang, Zemin
 ; TITLE OF INVENTION: SECRETED AND TRANSMEMBRANE POLYPEPTIDES AND NUCLEIC
 ; FILE REFERENCE: P3330R1C154
 ; CURRENT APPLICATION NUMBER: US/10/137,865
 ; CURRENT FILING DATE: 2002-05-03
 ; Prior Application removed - See Palm or File Wrapper
 ; NUMBER OF SEQ ID NOS: 550
 ; SEQ ID NO 504
 ; LENGTH: 163
 ; TYPE: PRT
 ; ORGANISM: Homo Sapien
 US-10-137-865-504

Query Match 1.6%; Score 8; DB 9; Length 163;
 Best Local Similarity 100.0%; Pred.No. 7.9;
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 81 LYSUTETA 88
 |||||
 Db 136 LYSUTETA 143

RESULT 11
US-10-140-474-504
; Sequence 504, Application US/10140474
; Publication No. US20030032156A1
; GENERAL INFORMATION:
; APPLICANT: Baker, Kevin P.
; APPLICANT: Beresini, Maureen
; APPLICANT: DeForge, Laura
; APPLICANT: Desnoyers, Luc
; APPLICANT: Filvaroff, Ellen
; APPLICANT: Gao, Wei-Qiang
; APPLICANT: Gerritsen, Mary E.
; APPLICANT: Goddard, Audrey
; APPLICANT: Godowski, Paul J.
; APPLICANT: Gurney, Austin L.
; APPLICANT: Sherwood, Steven
; APPLICANT: Smith, Victoria
; APPLICANT: Stewart, Timothy A.
; APPLICANT: Tumas, Daniel
; APPLICANT: Watanabe, Colin K
; APPLICANT: Wood, William
; APPLICANT: Zhang, Zemin
; TITLE OF INVENTION: SECRETED AND TRANSMEMBRANE POLYPEPTIDES AND NUCLEIC
; FILE REFERENCE: P3330RIC162
; CURRENT APPLICATION NUMBER: US/10/140,474
; CURRENT FILING DATE: 2002-05-06
; Prior Application removed - See Palm or File Wrapper
; NUMBER OF SEQ ID NOS: 550
; SEQ ID NO 504
; LENGTH: 163
; TYPE: PRT
; ORGANISM: Homo Sapien
US-10-140-474-504

Query Match 1.6%; Score 8; DB 9; Length 163;
Best Local Similarity 100.0%; Pred. No. 7.9;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 81 LYSLTETA 88
| | | | | | | |
Db 136 LYSLTETA 143

RESULT 12
US-10-142-431-504
; Sequence 504, Application US/10142431
; Publication No. US20030036179A1
; GENERAL INFORMATION:
; APPLICANT: Baker, Kevin P.
; APPLICANT: Beresini, Maureen
; APPLICANT: DeForge, Laura
; APPLICANT: Desnoyers, Luc
; APPLICANT: Filvaroff, Ellen
; APPLICANT: Gao, Wei-Qiang
; APPLICANT: Gerritsen, Mary E.
; APPLICANT: Goddard, Audrey
; APPLICANT: Godowski, Paul J.
; APPLICANT: Gurney, Austin L.
; APPLICANT: Sherwood, Steven
; APPLICANT: Smith, Victoria
; APPLICANT: Stewart, Timothy A.
; APPLICANT: Tumas, Daniel
; APPLICANT: Watanabe, Colin K
; APPLICANT: Wood, William
; APPLICANT: Zhang, Zemin
; TITLE OF INVENTION: SECRETED AND TRANSMEMBRANE POLYPEPTIDES AND NUCLEIC
; FILE REFERENCE: P3330RIC251
; CURRENT APPLICATION NUMBER: US/10/142,431
; CURRENT FILING DATE: 2002-05-10

; Prior Application removed - See File Wrapper or Palm
; NUMBER OF SEQ ID NOS: 550
; SEQ ID NO 504
; LENGTH: 163
; TYPE: PRT
; ORGANISM: Homo Sapien
US-10-142-431-504

Query Match 1.6%; Score 8; DB 9; Length 163;
Best Local Similarity 100.0%; Pred. No. 7.9;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 81 LYSLTETA 88
| | | | | | | |
Db 136 LYSLTETA 143

RESULT 13
US-10-143-114-504
; Sequence 504, Application US/10143114
; Publication No. US20030036180A1
; GENERAL INFORMATION:
; APPLICANT: Baker, Kevin P.
; APPLICANT: Beresini, Maureen
; APPLICANT: DeForge, Laura
; APPLICANT: Desnoyers, Luc
; APPLICANT: Filvaroff, Ellen
; APPLICANT: Gao, Wei-Qiang
; APPLICANT: Gerritsen, Mary E.
; APPLICANT: Goddard, Audrey
; APPLICANT: Godowski, Paul J.
; APPLICANT: Gurney, Austin L.
; APPLICANT: Sherwood, Steven
; APPLICANT: Smith, Victoria
; APPLICANT: Stewart, Timothy A.
; APPLICANT: Tumas, Daniel
; APPLICANT: Watanabe, Colin K
; APPLICANT: Wood, William
; APPLICANT: Zhang, Zemin
; TITLE OF INVENTION: SECRETED AND TRANSMEMBRANE POLYPEPTIDES AND NUCLEIC
; FILE REFERENCE: P3330RIC211
; CURRENT APPLICATION NUMBER: US/10/143,114
; CURRENT FILING DATE: 2002-05-09
; Prior Application removed - See Palm or File Wrapper
; NUMBER OF SEQ ID NOS: 550
; SEQ ID NO 504
; LENGTH: 163
; TYPE: PRT
; ORGANISM: Homo Sapien
US-10-143-114-504

Query Match 1.6%; Score 8; DB 9; Length 163;
Best Local Similarity 100.0%; Pred. No. 7.9;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 81 LYSLTETA 88
| | | | | | | |
Db 136 LYSLTETA 143

RESULT 14
US-10-230-163-184
; Sequence 184, Application US/10230163
; Publication No. US20030036635A1
; GENERAL INFORMATION:
; APPLICANT: Baker, Kevin P.
; APPLICANT: Desnoyers, Luc
; APPLICANT: Gerritsen, Mary
; APPLICANT: Goddard, Audrey
; APPLICANT: Godowski, Paul J.
; APPLICANT: Grimaldi, J. Christopher
; APPLICANT: Gurney, Austin L.

APPLICANT: Smith, Victoria
APPLICANT: Stephan, Jean-Philippe F.
APPLICANT: Watanabe, Colin L.
APPLICANT: Wood, William I.
TITLE OF INVENTION: SECRETED AND TRANSMEMBRANE POLYPEPTIDES AND NUCLEIC
FILE REFERENCE: P3530PTC96
CURRENT APPLICATION NUMBER: US/10/230,163
CURRENT FILING DATE: 2002-08-28
PRIOR APPLICATION NUMBER: 10/119,480
PRIOR FILING DATE: 2002-04-09
PRIOR APPLICATION NUMBER: 60/0959113
PRIOR FILING DATE: 1997-09-17
PRIOR APPLICATION NUMBER: 60/062287
PRIOR FILING DATE: 1997-10-17
PRIOR APPLICATION NUMBER: 60/063549
PRIOR FILING DATE: 1997-10-28
PRIOR APPLICATION NUMBER: 60/064103
PRIOR FILING DATE: 1997-10-31
PRIOR APPLICATION NUMBER: 60/069873
PRIOR FILING DATE: 1997-12-17
PRIOR APPLICATION NUMBER: 60/078910
PRIOR FILING DATE: 1998-03-20
PRIOR APPLICATION NUMBER: 60/079294
PRIOR FILING DATE: 1998-03-25
PRIOR APPLICATION NUMBER: 60/079656
PRIOR FILING DATE: 1998-03-26
PRIOR APPLICATION NUMBER: 60/079728
PRIOR FILING DATE: 1998-03-27
PRIOR APPLICATION NUMBER: 60/081819
PRIOR FILING DATE: 1998-04-15
PRIOR APPLICATION NUMBER: 60/081955
PRIOR FILING DATE: 1998-04-15
PRIOR APPLICATION NUMBER: 60/082804
PRIOR FILING DATE: 1998-04-22
PRIOR APPLICATION NUMBER: 60/084441
PRIOR FILING DATE: 1998-05-06
PRIOR APPLICATION NUMBER: 60/085323
PRIOR FILING DATE: 1998-05-13
PRIOR APPLICATION NUMBER: 60/085579
PRIOR FILING DATE: 1998-05-15
PRIOR APPLICATION NUMBER: 60/086392
PRIOR FILING DATE: 1998-05-22
PRIOR APPLICATION NUMBER: 60/089532
PRIOR FILING DATE: 1998-06-17
PRIOR APPLICATION NUMBER: 60/089538
PRIOR FILING DATE: 1998-06-17
PRIOR APPLICATION NUMBER: 60/089905
PRIOR FILING DATE: 1998-06-18
PRIOR APPLICATION NUMBER: 60/090472
PRIOR FILING DATE: 1998-06-24
PRIOR APPLICATION NUMBER: 60/090557
PRIOR FILING DATE: 1998-06-24
PRIOR APPLICATION NUMBER: 60/090691
PRIOR FILING DATE: 1998-06-25
PRIOR APPLICATION NUMBER: 60/090695
PRIOR FILING DATE: 1998-06-25
PRIOR APPLICATION NUMBER: 60/091982
PRIOR FILING DATE: 1998-07-07
PRIOR APPLICATION NUMBER: 60/095302
PRIOR FILING DATE: 1998-08-04
PRIOR APPLICATION NUMBER: 60/095318
PRIOR FILING DATE: 1998-08-04
PRIOR APPLICATION NUMBER: 60/095916
PRIOR FILING DATE: 1998-08-10
PRIOR APPLICATION NUMBER: 60/096146
PRIOR FILING DATE: 1998-08-11
PRIOR APPLICATION NUMBER: 60/096791
PRIOR FILING DATE: 1998-08-17
PRIOR APPLICATION NUMBER: 60/097986
PRIOR FILING DATE: 1998-08-26
PRIOR APPLICATION NUMBER: 60/098544
PRIOR FILING DATE: 1998-08-31
PRIOR APPLICATION NUMBER: 60/099596
PRIOR FILING DATE: 1998-09-09
PRIOR APPLICATION NUMBER: 60/099598
PRIOR FILING DATE: 1998-09-09
PRIOR APPLICATION NUMBER: 60/099803
PRIOR FILING DATE: 1998-09-10
PRIOR APPLICATION NUMBER: 60/099811
PRIOR FILING DATE: 1998-09-10
PRIOR APPLICATION NUMBER: 60/099812
PRIOR FILING DATE: 1998-09-10
PRIOR APPLICATION NUMBER: 60/099816
PRIOR FILING DATE: 1998-09-10
PRIOR APPLICATION NUMBER: 60/100038
PRIOR FILING DATE: 1998-09-11
PRIOR APPLICATION NUMBER: 60/100385
PRIOR FILING DATE: 1998-09-15
PRIOR APPLICATION NUMBER: 60/100390
PRIOR FILING DATE: 1998-09-15
PRIOR APPLICATION NUMBER: 60/100627
PRIOR FILING DATE: 1998-09-16
PRIOR APPLICATION NUMBER: 60/100848
PRIOR FILING DATE: 1998-09-18
PRIOR APPLICATION NUMBER: 60/100919
PRIOR FILING DATE: 1998-09-17
PRIOR APPLICATION NUMBER: 60/101477
PRIOR FILING DATE: 1998-09-23
PRIOR APPLICATION NUMBER: 60/101738
PRIOR FILING DATE: 1998-09-24
PRIOR APPLICATION NUMBER: 60/101741
PRIOR FILING DATE: 1998-09-24
PRIOR APPLICATION NUMBER: 60/101786
PRIOR FILING DATE: 1998-09-25
PRIOR APPLICATION NUMBER: 60/101916
PRIOR FILING DATE: 1998-09-24
PRIOR APPLICATION NUMBER: 60/101922
PRIOR FILING DATE: 1998-09-24
PRIOR APPLICATION NUMBER: 60/106178
PRIOR FILING DATE: 1998-10-28
PRIOR APPLICATION NUMBER: 60/106248
PRIOR FILING DATE: 1998-10-29
PRIOR APPLICATION NUMBER: 60/106464
PRIOR FILING DATE: 1998-10-30
PRIOR APPLICATION NUMBER: 60/106905
PRIOR FILING DATE: 1998-11-03
PRIOR APPLICATION NUMBER: 60/108787
PRIOR FILING DATE: 1998-11-17
PRIOR APPLICATION NUMBER: 60/108801
PRIOR FILING DATE: 1998-11-17
PRIOR APPLICATION NUMBER: 60/108849
PRIOR FILING DATE: 1998-11-18
PRIOR APPLICATION NUMBER: 60/112422
PRIOR FILING DATE: 1998-12-15
PRIOR APPLICATION NUMBER: 60/113296
PRIOR FILING DATE: 1998-12-22
PRIOR APPLICATION NUMBER: 60/113605
PRIOR FILING DATE: 1998-12-23
PRIOR APPLICATION NUMBER: 60/113621
PRIOR FILING DATE: 1998-12-23
PRIOR APPLICATION NUMBER: 60/115558
PRIOR FILING DATE: 1999-01-12
PRIOR APPLICATION NUMBER: 60/115565
PRIOR FILING DATE: 1999-01-12
PRIOR APPLICATION NUMBER: 60/115733
PRIOR FILING DATE: 1999-01-12
PRIOR APPLICATION NUMBER: 60/119549
PRIOR FILING DATE: 1999-02-10
PRIOR APPLICATION NUMBER: 60/123618
PRIOR FILING DATE: 1999-03-10
PRIOR APPLICATION NUMBER: 60/125259
PRIOR FILING DATE: 1999-03-19
PRIOR APPLICATION NUMBER: 60/125775
PRIOR FILING DATE: 1999-03-23
PRIOR APPLICATION NUMBER: 60/126773

```

; PRIOR FILING DATE: 1999-03-29
; PRIOR APPLICATION NUMBER: 60/127887
; PRIOR FILING DATE: 1999-04-05
; PRIOR APPLICATION NUMBER: 60/130232
; PRIOR FILING DATE: 1999-04-21
; PRIOR APPLICATION NUMBER: 60/131022
; PRIOR FILING DATE: 1999-04-26
; PRIOR APPLICATION NUMBER: 60/131270
; PRIOR FILING DATE: 1999-04-27
; PRIOR APPLICATION NUMBER: 60/131291
; PRIOR FILING DATE: 1999-04-27
; PRIOR APPLICATION NUMBER: 60/131445
; PRIOR FILING DATE: 1999-04-28
; PRIOR APPLICATION NUMBER: 60/134287
; PRIOR FILING DATE: 1999-05-14
; PRIOR APPLICATION NUMBER: 60/140650
; PRIOR FILING DATE: 1999-06-22
; PRIOR APPLICATION NUMBER: 60/140723
; PRIOR FILING DATE: 1999-06-22
; PRIOR APPLICATION NUMBER: 60/141037
; PRIOR FILING DATE: 1999-06-23
; PRIOR APPLICATION NUMBER: 60/144758
; PRIOR FILING DATE: 1999-07-20
; PRIOR APPLICATION NUMBER: 60/145698
; PRIOR FILING DATE: 1999-07-26
; PRIOR APPLICATION NUMBER: 60/146222
; PRIOR FILING DATE: 1999-07-28
; PRIOR APPLICATION NUMBER: 60/146963
; PRIOR FILING DATE: 1999-08-03
; PRIOR APPLICATION NUMBER: 60/149320
; PRIOR FILING DATE: 1999-08-17
; PRIOR APPLICATION NUMBER: 60/149638
; PRIOR FILING DATE: 1999-08-17
; PRIOR APPLICATION NUMBER: 60/151733
; PRIOR FILING DATE: 1999-08-31
; PRIOR APPLICATION NUMBER: 60/164418
; PRIOR FILING DATE: 1999-11-09
; PRIOR APPLICATION NUMBER: 60/166361
; PRIOR FILING DATE: 1999-11-16
; PRIOR APPLICATION NUMBER: 60/169445
; PRIOR FILING DATE: 1999-12-07
; PRIOR APPLICATION NUMBER: 60/169495
; PRIOR FILING DATE: 1999-12-07
; PRIOR APPLICATION NUMBER: 60/169835

```

```

Query Match      1.6%; Score 8; DB 9; Length 163;
Best Local Similarity 100.0%; Pred. No. 7.9;
Matches      8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

```

Qy      81 LYSLETETA 88
Db      136 LYSLETETA 143

```

RESULT 15

```

US-10-140-002-504
; Sequence 504, Application US/10140002
; Publication No. US20030037623A1
; GENERAL INFORMATION:
; APPLICANT: Baker, Kevin P.
; APPLICANT: Beresini, Maureen
; APPLICANT: DeForge, Laura
; APPLICANT: Desnoyers, Luc
; APPLICANT: Filvaroff, Ellen
; APPLICANT: Gao, Wei-Qiang
; APPLICANT: Gerritsen, Mary E.
; APPLICANT: Goddard, Audrey
; APPLICANT: Godowski, Paul J.
; APPLICANT: Gurney, Austin L.
; APPLICANT: Sherwood, Steven
; APPLICANT: Smith, Victoria
; APPLICANT: Stewart, Timothy A.
; APPLICANT: Tumas, Daniel

```

```

; APPLICANT: Watanabe, Colin K
; APPLICANT: Wood, William
; APPLICANT: Zhang, Zemin
; TITLE OF INVENTION: SECRETED AND TRANSMEMBRANE POLYPEPTIDES AND NUCLEIC
; FILE REFERENCE: P3330RIC59
; CURRENT APPLICATION NUMBER: US/10/140.002
; CURRENT FILING DATE: 2002-05-06
; Prior Application removed - See Palm or File Wrapper
; NUMBER OF SEQ ID NOS: 550
; SEQ ID NO 504
; LENGTH: 163
; TYPE: PRT
; ORGANISM: Homo Sapien
US-10-140-002-504

```

```

Query Match      1.6%; Score 8; DB 9; Length 163;
Best Local Similarity 100.0%; Pred. No. 7.9;
Matches      8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

```

Qy      81 LYSLETETA 88
Db      136 LYSLETETA 143

```

```

Search completed: May 8, 2003, 16:41:48
Job time : 34 secs

```

THIS PAGE BLANK (USPTO)